

A Quasi-Experimental Study of Formative Peer Assessment in an EFL Writing Classroom



**A Quasi-Experimental Study of Formative Peer Assessment in an EFL
Writing Classroom**

Chia-Lin Kuo

**Thesis Submitted for the Degree of
Doctor of Philosophy**

Integrated PHD in Applied Linguistics and TESOL

**NEWCASTLE UNIVERSITY
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
SCHOOL OF EDUCATION, COMMUNICATION AND
LANGUAGE SCIENCES**

April 2015



**NEWCASTLE UNIVERSITY
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
SCHOOOL OF EDUCATION, COMMUNICATION AND
LANGUAGE SCIENCES**

DECLARATION

I hereby certify that this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or currently submitted for any other degree at the University of Newcastle or other institutions.

Name: Chia-Lin Kuo

Signature:

Date: 12. 04. 2015

Acknowledgements

I would like to express my greatest gratitude to my supervisory team. I would also like to thank my family and friends for their unconditional support and encouragement.

Abstract

The benefits of using formative peer assessment have been broadly recognized by educators. However, in the context of English as a Foreign Language (EFL) teaching, research into the effectiveness of peer assessment has produced mixed results. One major focus of research is how training can influence the effectiveness of peer assessment. A number of case studies have reported positive outcomes using particular training methods, but there is relatively little research into the application of such training methods and their effects in ordinary EFL classes. This thesis reports a study of training in peer assessment in the context of a typical EFL class. The details of the training are explored in some depth, including evidence of its effects on the nature and quality of peer assessment and the reasons for those effects.

A quasi-experiment was designed to investigate the effectiveness of the extensively discussed ‘Step Training’ proposed by Min. Data were collected in the form of students’ written feedback, essays, video recording, questionnaires and interviews. The nature of students’ feedback and the quality of their essays were explored. Classroom interactions were analysed and interpreted using socio-cultural theory. Analysis of the questionnaires revealed students’ attitude towards the training and perception of their capabilities in working on peer assessment. Student interview data was analysed thematically.

The findings provide a complex picture of the peer assessment training. First, the ‘Step training’ appeared to lead to improvements in the quality of both essays and peer feedback. Secondly, analysis of the classroom discourse suggested that socio-cultural factors had both positive and negative effects on the students’ learning. Moreover, ANCOVA analysis of psychological features such as attitudes towards and perceptions of the peer assessment training suggested that the students were less influenced psychologically by the training. Finally, analysis of interviews identified students’ concerns about the design of the training course.

In conclusion, it is suggested that for a typical EFL writing class, the effectiveness of a rigid training method such as ‘Step training’ should be reconsidered to take into account the influence of socio-cultural factors in classroom interaction. Rather than relying just on students’ end products to monitor the effectiveness of training, socio-cultural interaction should also be examined, as this is important in developing an identity as an effective peer assessor. Suggestions for improving the design of peer assessment training are provided.

Table of Contents

DECLARATION	I
ACKNOWLEDGEMENTS	II
ABSTRACT	III
FIGURES	VII
TABLES	VII
EXCERPTS	VIII
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Rationale of the Study	2
1.3 Objective of the Study	3
1.4 Research Questions	4
1.5 Value of the Study	5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Peer Assessment	6
2.2.1 Definition of Formative Peer Assessment	7
2.2.2 Theoretical Review of Peer Assessment	7
2.2.3 Benefits of Peer Assessment	11
2.2.4 Difficulties of Peer Assessment in L2 Writing	12
2.2.5 Influence of Students' Psychological Characteristics on Peer Assessment Performance	15
2.3 Review of Training for Peer Assessment	18
2.3.1 Definition of 'Effective' Peer Feedback	18
2.3.2 Training Methods for Effective Peer Assessment	19
2.4 Discourse and Interaction	22
2.4.1 Regulative Discourse and Instructional Discourse	22
2.4.2 Classification and Framing	22
2.5 Teacher's Role in Classroom Peer Assessment	26
2.6 Research Gaps	27
2.7 Summary	27
CHAPTER 3 METHODOLOGY	29
3.1 Aims	29
3.2 Critique of Methods Used in Previous Work	31
3.3 Methodological Basis	33
3.3.1 Rationale for Research Methods	33
3.3.2 Nature of Research	34
3.4 Research Design	34
3.4.1 Overview of Data Collection Procedures	35
3.4.2 Preparation Activities	36
3.4.3 Course Design	38
3.4.4 Peer Assessment Tasks	40

3.5 Participants and Sampling.....	40
3.6 Justification of Data Collection Tools.....	43
3.6.1 Students' Written Peer Feedback.....	43
3.6.2 Students' Argumentative Writing Assignments	43
3.6.3 Questionnaire	44
3.6.4 Interviews.....	45
3.6.5 Video Recording	45
3.7 Data Analysis	46
3.7.1 Analysis of Students' Written Feedback.....	46
3.7.2 Analysis of Students' Argumentative Writing Assignments	48
3.7.3 Analysis of Questionnaire Results	49
3.7.4 Recordings of Interviews	49
3.7.5 Analysis of Video Recordings	50
3.8 Reliability and Validity	51
3.9 Ethical Integrity.....	52
CHAPTER 4 RESULTS	53
4.1 Introduction	53
4.2 Background Analysis	54
4.3 Step Usage and the Quality of Students' Work	54
4.3.1 Statistical Methods	55
4.3.2 The 'Use of Steps' in Students' Formative Peer Feedback.....	55
4.3.3 The Quality of Students' Formative Peer Feedback	58
4.3.4 The Quality of Students' Written Assignments	61
4.3.5 Findings.....	63
4.4 Students' Opinions, Attitudes and Perceptions.....	64
4.4.1 Students' Perception of their Ability to Produce Peer Feedback.....	65
4.4.2 Students' Perception of Working on Peer Feedback.....	66
4.4.3 Students' Perceptions of Writing	68
4.4.4 Students' Attitudes towards Peer Assessment Training	69
4.4.5 Students' Opinions of Time Allocated to Peer Assessment	70
4.4.6 Students' Opinions of Peer Collaboration in Peer Feedback Activity.....	71
4.4.7 Students' Attitudes towards the Teacher's Role in Assessment.....	72
4.4.8 Students' Evaluation of Motivation	73
4.4.9 Relationship between Psychological Factors and Peer Assessment Performance	74
4.5 Analysis of Classroom Interaction	75
4.5.1 Observing Two Discourses	75
4.5.2 Classification and Framing in the Discourses.....	76
4.5.3 Teacher's Role and Techniques	88
4.5.4 From Convergent to Divergent Assessment	91
4.5.5 Construction of the Student's Identity in Peer Assessment Training	94
4.6 Summary	96
4.6.1 Peer Feedback and Writing	96
4.6.2 Students' Attitudes, Perceptions and Opinions.....	97
4.6.3 Classroom Interaction	97
CHAPTER 5 DISCUSSION	99
5.1 Results for the 'Step' Training for Peer Assessment	99
5.1.1 Positive Evidence from the Experimental Group	99
5.1.2 Weaknesses in the Measurement of "Steps"	99
5.1.3 Theoretical Analysis of the Steps.....	100

5. 2 Students' Perceptions and Attitudes.....	101
5.2.1 Types of Beliefs	101
5.2.2 Measurement of Psychological Factors	101
5.2.3 Students' Perceptions and Peer Assessment Quality	102
5.3 Socio-Cultural Learning and Peer Assessment	105
5.3.1 Difficulties in Student Identity Construction during Step Training.....	106
5.3.2 Teacher's Role	113
CHAPTER 6 CONCLUSION.....	117
APPENDIX A TRAINING STEPS SUGGESTED BY MIN (2005).....	121
APPENDIX B PEER ASSESSMENT WORKSHEET	122
APPENDIX C LETTER OF CONSENT FOR THE PARTICIPANTS IN THIS STUDY	126
APPENDIX D STUDENT INTERVIEW QUESTIONS	127
APPENDIX E QUESTIONNAIRE QUESTIONS	129
APPENDIX F CODING FOR STUDENTS WRITTEN FEEDBACK.....	130
APPENDIX G CODING EXAMPLES FOR STUDENTS' ARGUMENTATIVE WRITING	131
APPENDIX H TRANSCRIPTION CONVENTIONS	133
BIBLIOGRAPHY	135

Figures

Figure 2-1 Theory of planned behaviour	16
Figure 2-2 Conceptual framework	28
Figure 4-1 Framing in the negotiation discourse	88
Figure 4-2 Formative assessment for step training	93

Tables

Table 2-1 Definitions of Min's (2005) four steps	18
Table 4- 1 Mean scores and independent <i>t</i> -test results for English language proficiency	54
Table 4-2 Mean scores and independent <i>t</i> -test result for step usage	57
Table 4-3 ANCOVA results comparing the use of steps by the two groups	57
Table 4-4 Mean scores and independent <i>t</i> -test results for the quality of peer feedback	58
Table 4-5 ANOVA results for the quality of peer feedback between tasks	58
Table 4-6 Paired <i>t</i> -test results for the quality of students' writing before and after training	62
Table 4-7 Mean scores and independent <i>t</i> -test result for the quality of students' writing	62
Table 4-8 ANCOVA results for the quality of students' writing	62
Table 4-9 Mean scores and independent <i>t</i> -test results for Q7	65
Table 4-10 Mean variations and paired <i>t</i> -test results for Q7	66
Table 4-11 Mean scores and independent <i>t</i> -test results for Q8, 9 and 10	67
Table 4-12 Mean variations and paired <i>t</i> -test results for Q8,9 and 10	67
Table 4-13 Mean scores and independent <i>t</i> -test results for Q16	69
Table 4-14 Mean scores and independent <i>t</i> -test results for Q6	69
Table 4-15 Mean variations and paired <i>t</i> -test results for Q6	69
Table 4-16 Mean scores and independent <i>t</i> -test results for Q14	70
Table 4-17 Mean scores and independent <i>t</i> -test results for Q11, 12 and 13	71
Table 4-18 Means difference and paired <i>t</i> -test results for Q11, 12 and 13	71
Table 4-19 Mean scores and independent <i>t</i> -test results for Q1 and 4	72
Table 4-20 Mean differences and paired <i>t</i> -test results for Q1 and 4	73
Table 4-21 Mean scores and independent <i>t</i> -test results for Q2 and 3	73
Table 4-22 Mean differences and paired <i>t</i> -test results for Q2 and 3	73
Table 4-23 ANCOVA psychological factors and peer assessment performance	74
Table 4-24 Definitions of framing and strengths	76
Table 4-25 Definitions of classification and strengths	77
Table 4-26 The techniques and teacher's role in 2 types of framing combination	91
Table 4-27 Convergent to divergent formative assessment for learning	94

Excerpts

Excerpt 1	78
Excerpt 2	81
Excerpt 3	83
Excerpt 4	85
Excerpt 5	87
Excerpt 6	95

Chapter 1 Introduction

1.1 Background

The use of peer assessment with learners is attracting the attention of increasing numbers of teachers (Sadler et.al. 2006) and researchers who are interested in exploring the potential benefits of this technique in a wide range of subjects (Sadler et.al. 2006). This interest has resulted in a growing number of publications reporting on both theoretical and practical aspects of this topic (Topping, 2009).

As a result of such studies, it is now widely accepted that peer assessment can benefit students in a number of ways (Bernstein, 1996), one of the most widely reported of which is its cognitive benefits. It has been reported that peer assessment can stimulate students' thinking (Nicol and Macfarlane, 2006; Hughes, 2009) when they debate opinions, discuss strategies and share perspectives. As they engage in such a process, students are also required to take an active learning role in which they can "reconceptualise their ideas in light of their peers' reactions" (Mendoca and Johnson 1994:746).

Students can also develop evaluative skills which help them to analyse and evaluate the feedback and revise their own work through peer assessment (Sadler, 1989; Leki, 1990b). These skills can later enable them to become self-monitoring and to more effectively self-regulate their learning (Topping, 2009).

The effectiveness of peer assessment has also been investigated in the context of teaching English as a foreign language (EFL). This is especially so in relation to writing, where negotiating the meaning of a text, and providing peer feedback, can then develop a sense of audience which can help students adapt their writing to suit the readers they are writing for. As a result, peer assessment can improve the quality of a student's writing (Topping et al, 2000).

Moreover, students can also develop "social and communication skills, negotiation and diplomacy, and useful transferable skills like giving and handling criticism, self-justification and assertion" (Topping, 2003:67). Peer assessment offers students the chance to test out their hypotheses about their language, especially for second language (L2) students (Liu and Hansen, 2002). They also have the opportunity to revise their

work in a supportive environment in which they can practise and apply their writing skills under less stressful conditions (Villamil and Guerrero, 1996; 1998).

Liu and Hansen (2002) identified the linguistic and practical benefits of peer assessment with L2 students. In terms of linguistic benefits, students were able to discover and negotiate meaning in peer assessment as they explored efficient ways to express meanings (Mendonca and Johnson, 1994; Nelson and Carson, 1998; Liu and Sadler, 2003).

More practically, peer assessment in a writing class is also considered to be flexible across different stages of the writing cycle, such as in multiple drafts, extensive revision, and pair/group work, to encourage improved discussion (Liu and Hansen, 2002; Hu, 2005).

Last but not least, the reduced workload in terms of reduced marking tasks for L2 students is a direct benefit to tutors of successful peer assessment (Topping et al, 2000; Loddington, 2008). Race (2001:17) pointed out that, for academics/tutors, peer assessment makes it easier to see “where the strengths and weaknesses of a cohort of students lie”. Over time, teachers can identify these common weaknesses and act upon them.

Learning to write in English has always been challenging for Taiwanese students. From the past teaching experiences of the present author, even average academic level college students still struggle to succeed in writing in English. As a result of being inspired by previous studies, the author was interested in exploring the potential benefits of using peer assessment with an intermediate proficiency level of English class. To develop students’ writing skills, peer assessment was introduced in writing courses as a pedagogical activity and as formative assessment to encourage more cooperative and in-depth learning.

1.2 Rationale of the Study

The application of peer assessment in classroom teaching has been widely researched. However, not many studies have focussed on the application of peer assessment within the context of English teaching/learning as a foreign language (EFL). Three particular relevant studies on this topic have been published. Zhu (1995) conducted a study with freshmen (n=169) in a university in the USA where the majority (n=144) were native speakers. In her study, she adopted a teacher-student-conference method in which the

teacher meets students in groups to discuss strategies for providing effective feedback on peer writing. Although this research was not performed in the EFL context, her method has influenced many subsequent studies in the EFL/ESL context.

Hu (2005) performed an empirical research study of peer assessment in an ESL context. She carried out a 3 year-long action research study into the most effective teaching method for training learners to use peer feedback. She designed 24 activities in which students provided peer feedback on writing and provided a number of technical recommendations for effective peer assessment.

Min (2005; 2006) wanted to improve the results of paired peer assessment by giving her students training. This included 4 hours of demonstration in how to give comments, and 2 teacher-student conference sessions. She concluded that this 'step' training design would significantly improve the quality and effectiveness of peer feedback and, therefore, could improve the effectiveness of peer assessment.

Zhu's (1995), Hu's (2005) and Min's (2005) work is based on the same simple idea: peer assessment results can be improved by improving the training methods. They all designed, tested and investigated routines and methods specifically aimed at training their students to give comments, which their results suggested were effective.

However, there appear to be a number of issues with their methods. First of all, the students in their samples were well qualified (or native speakers) and so might not be representative of the majority of L2 students. Secondly, while their recommendations are based on empirical investigation and quantitative analysis of their results, there is little attempt to provide a theoretical justification or framework for the effectiveness of their training methods.

In addition, when the present author tried to duplicate the success they reported for their training methods, the difficulties encountered led the author to reconsider the effectiveness of methods such as those proposed by Zhu (1995), Hu (2005) and Min (2005)

1.3 Objective of the Study

The purpose of this study is therefore to investigate the effectiveness of peer assessment training schemes when used with what the author considers to be typical, average L2 students, and to develop a theoretical framework for peer assessment. In pursuit of this

goal, a quasi-experiment of peer assessment training was conducted which was designed based on Min's (2005) peer assessment training system with a class of EFL college students.

Unlike in previous research, the present research was carried out with lower intermediate level (as opposed to advanced) students. The study outcome were then interpreted from two theoretical points of view, and the results included qualitative as well as quantitative measures.

One of the study's two theoretical viewpoints is the theory of planned behaviour. The author expects that the measurement of the perceptions and opinions of students will provide evidence of the impact of the training and how the peer assessment training was received. The second viewpoint is the socio-cultural theory of learning, which forms the basis for the theoretical framework used in this research. In investigating in detail the interaction in the classroom, this theory will be used to provide explanations of why some kinds of interaction are more effective than others, and of the overall effectiveness of the training.

1.4 Research Questions

The main research question in this study is:

How does the nature of pedagogic discourse which the teacher provides in the step training impact on students' peer feedback?

The first of two sub-questions examines the effects of the training and is split into four further sub-questions:

1. What is the nature of peer feedback provided in both groups?
 - a. To what extent does the peer feedback follow Min's four steps as a result of Min's training method?
 - b. What is the quality of students' argumentative writing?
 - c. What are students' attitudes towards and perceptions of peer assessment before and after the training?

The second sub-question examines the pedagogical discourse of the training and is split into two further sub-questions:

2. How does the teacher model and negotiate in order to shape students' peer feedback?

- a. Based on Bernstein's (1996) concepts of classification and framing, how explicit is the modelling and negotiation in providing guidance to the students?
- b. How does the teacher engage students in the discourse of modelling and negotiation?

1.5 Value of the Study

This study explores the use of training and examines its effects on EFL formative peer assessment and students' attitudes in a Taiwanese college, which the author of the present study considers to be representative of an average level college (see section 3.4). This study makes an original contribution to the field by carrying out detailed mixed-method research on peer assessment training in a typical college-level EFL writing teaching context. The research includes a detailed analysis of interactions in the classroom using discourse analysis, which has not been provided in previous studies. The research provides a theoretical explanation of peer assessment training in this context, and the results can be used to inform recommendations toward effective peer assessment training based on socio-cultural theory.

Chapter 2 Literature Review

2.1 Introduction

Chapter 1 has provided a broad overview of the historical and contemporary development of peer assessment in educational applications and identified the importance of formative peer assessment in higher education and in second language teaching.

In this chapter, the definition, underpinning theories and application of peer assessment are reviewed so as to explore its effects in EFL classrooms. This is followed by a detailed review of past research on training programmes associated with the concept of formative assessment, which are intended to help students to become peer reviewers.

To further understand the students' learning process during peer assessment training, the socio-cultural theory of formative assessment is reviewed in detail, accompanied by discussions based on previous research results and theories. This informs the structure of the main theoretical framework of the study and helps to identify gaps in existing research and to inform the research questions asked.

2.2 Peer Assessment

The development of modern society has brought about new challenges for the traditional functions of school teaching, and how students should be 'told what to do and how to do it' (Blisshen, 1969). Students are expected to have the ability and motivation to pursue self-improvement and lifelong learning, and in addition concepts of personalised learning and learner-centred culture increasingly popular.

Consequently, innovations in classroom teaching and assessment are driven by these demands. In response, peer and self-assessment are among the most widely reviewed and practiced innovations in the classroom.

As early as the 1950s, research into peer assessment investigated its reliability as an alternative method to traditional assessment (Kubany, 1957). However, contemporary research and practice in peer assessment has gone far beyond its original boundaries. In the 1990s, the thinking behind peer assessment began to build momentum and evidence of improved student achievement was reported (Black and Wiliam, 1998). Teachers started to recognise the benefits of getting students involved in self and peer assessment.

The concept of involving students in the assessment of their own learning started to gain more acceptance (Deakin et al, 2007).

2.2.1 Definition of Formative Peer Assessment

Topping et al. (2000) defined peer assessment as “an arrangement for peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others of similar status.” Peer assessment can be applied in different forms as distinguished by Kane and Lawler (1978): peer ranking; peer nomination and peer rating. Peer evaluation is also an alternative term for peer assessment (Weaver and Cotrell, 1989).

In this work, the discussion focuses on the design and application of formative peer assessment. Topping et al. (2000) pointed out that, in formative peer assessment, “peers are a source of formative assessment with an intention to help each other’s learning, identify their strength and weakness, target areas for remedial action and develop a better product or performance.”

In the literature, formative peer assessment has also been called peer review (e.g. Min, 2005), peer response (e.g. Zhu, 1995), peer tutoring (e.g. Falchikov, 2001) and peer evaluation (e.g. Kwok, 2008). Peer feedback is the essential part of peer assessment or review. It is a method of exchanging information between individuals or groups during peer assessment. Giving and reviewing peer feedback is the major part of peer assessment, and so the term ‘peer feedback’ is frequently informally used as the synonym for peer assessment.

In the following section, the term ‘peer feedback’ is used to refer to the written comments produced by students. ‘Peer feedback activity’ refers to the students’ collaboration resulting in peer feedback.

2.2.2 Theoretical Review of Peer Assessment

Ever since the first appearance of the term peer assessment in the literature (Kubany, 1957), it has been widely used and extensively reviewed in educational research. Various studies have collated a wide range of evidence on different aspects of peer assessment applications. However, peer assessment has been under-theorized until recently (Falchikov, 2001; Topping 2009). In both Topping and Falchikovs’ attempts to

theorize peer assessment, Piaget and Vygotsky's theories were referenced as the foundation of their work.

The cognitive theory encompassed the Piagetian school of thought and Vygotsky's description of scaffolding and the Zone of Proximal Development (ZPD), where cognitive development was a result of social interaction in which an individual was supported and guided by a more competent other within the ZPDs of both parties. The ZPD refers to the space between the person's actual level of development and their potential level of development. This theory influenced several later authors such as Liu and Hansen (2002): "...higher cognitive processes are hypothesized to emerge as a result of interaction, resulting in the individual's independent completion of the task, with the language use within the interaction serving as the critical device for mediating cognitive development".

Social psychology and theories relating to personal and professional development are also referenced frequently. In the work of Falchikov (2001), social-psychological theoretical perspectives are suggested (c.f. Falchikov, 2001: 106) which concern the behaviour and social skills of the writer and reviewer relating to their roles and relationships in the process of peer assessment. Falchikov claimed that by considering social-psychological theoretical perspectives, potential problems can be predicted and potential solutions suggested. Falchikov (2001) found that most social-psychological theoretical perspectives "endorse the importance of training" for peer assessment. For example, one problem referred to by social-psychologists as 'role conflict' could lead to difficulties arising during peer assessment as a result of the negative influence of an authority figure. The authority figure in this context denotes the person who is superior to both the tutor and student. The negative influence of an authority figure denotes the general disturbance or destructive effect on peer assessment as a result of interference by the authority figure. For example, during role conflict, the teacher / student may tend to appeal to the authority figure so that the process will be disrupted. Training prior to peer assessment is understood to be effective in addressing the negative influence on authority figure can have in assessment practice, and that such training can stimulate better relations in peer assessment (Mangelsdorf and Schlumberger, 1992).

As in Topping's (2009) work, theories related to personal and professional development are also referenced in Falchikov's theory. With peer assessment, students' understanding of the function of writing is essential. The development of their discipline

and motivation can improve their own writing practice in later revisions, in which they consider and respond to the requests of their audience (Liu and Hansen, 2002; Hu, 2005; Min, 2005).

Unlike Falchikov's reliance on a specific theory, Topping (2009) created a synthesis of these theories and composed a single theoretical model (Topping, 2009: 64, Figure 4.1). The model provided a systematic and holistic view of the practice cycle of peer assessment and included a wide range of influencing factors.

The theories above describe peer assessment in general applications. For peer assessment in a L2 writing practice context, a comprehensive theoretical review of relevant theory was conducted by Liu and Hansen (2002). They suggested that peer assessment study in the L2 context could take four theoretical stances:

- Process writing theory
- Collaborative learning theory
- Vygotsky's ZPD
- Interaction and second language acquisition

These stances support the use of peer assessment in the L2 writing classroom from both cognitive and psycholinguistic perspectives:

1. Process writing theory emerged around the late 1960s and early 1970s in response to the traditional product views of writing from around the late 1960s and early 1970s in L1 writing. The product view of writing focused on form over meaning and on the finished text. Acknowledging that writing in the real world is a dynamic, nonlinear and recursive process (Liu and Hansen, 2002), the theory focused on the process of writing.

The pedagogical use of this approach in students' learning to write engages them in a range of writing activities. A typical course following the process approach to writing encourages writers to brainstorm, outline, draft (focusing on meaning), rewrite (focusing on organization and meaning), and edit (focusing on style and grammar).

Peer assessment could support the process of L2 writing instruction with a focus on drafting and revision (Leki, 1990b; Mangelsdorf and Schlumberger, 1992) with students receiving peer feedback across various drafts.

2. Vygotsky's ZPD. Similar to the L1 context, research studies in the L2 context also commonly employ Vygotsky's theoretical framework to investigate interaction in group work and L2 writing (Villamil and Guerrero, 1996). These research studies aimed to examine how the application of peer assessment applied in second language writing classrooms may influence students' language learning.
3. The collaborative learning theory was heavily influenced by Vygotsky. One of its main claims is that learning, as well as knowledge, is socially constructed. Therefore, for learning and practice in writing, group work is essential in the classroom to allow students to learn through social interaction since knowledge and skills are developed through such interaction.

Based on this assumption, students' knowledge and skills are constructed through the negotiation of meaning. As stated by Bruffee (1984: 644), collaborative writing by 'pooling' resources between group members can help students to complete tasks they may not otherwise complete individually. The dialogue and interaction within the peer group is also beneficial to L2 writers. Students have increased opportunities to review and apply their knowledge of the second language in writing (Hirvela, 1999; Min, 2005).

4. Interaction and second language acquisition researchers (Varonis and Gass, 1985; Dobao, 2012) also suggested that interaction within groups allows second language skills to develop. The comprehensible input gained through interactional adjustments such as through the negotiation of meaning can modify the learning output (Long, 1996).

Long and Porter (1985) found that L2 students have more opportunities to practice and use language in two-way communication and a wider range of language functions was utilized for authentic purposes. With increased practice and authentic use of language, it was found that students made fewer corrections of errors but carried out more negotiation of meaning. These benefits indicate the advantage of adopting group work for second language acquisition (Long and Porter, 1985).

2.2.3 Benefits of Peer Assessment

A number of researchers have highlighted the potential benefits that peer assessment can have on both language learning and learner development. Firstly, it was reported that peer assessment could stimulate students' thinking (Nicola and Macfarlane, 2006; Hughes, 2009). In the process of negotiation, the students demonstrated and reinforced their knowledge about what they had learned and shared. Sharing perspectives and strategies in the process of negotiation enables them to "reconceptualise their ideas in light of their peers' reactions" (Mendoca and Johnson, 1994:746).

Secondly, the students were offered the opportunity to test out their hypotheses about their knowledge (Liu and Hansen, 2002). They could receive their peers' scaffolding and have the opportunity to revise their work in a supportive environment (Villamil and De Guerrero, 1996; 1998; Jacobs et al., 1998; Guerrero and Villamil, 2000; Cotterall and Cohen, 2003).

Thirdly, engaging in such a process helps develop their "social and communication skills, negotiation and diplomacy, and build useful transferable skills such as giving and handling criticism, self-justification and assertion" (Topping, 2003: 67).

Finally, students could also develop evaluative skills to help them analyse, evaluate feedback on and revise their work (Sadler, 1989; Leki, 1990a). These skills later enable them to become self-monitoring and self-regulative on their learning (Guerrero and Villamil, 1994; Nelson and Carson, 1998; Liu and Sadler, 2003; Topping, 2009), thereby becoming more autonomous learners and reducing their dependence on teachers (Tsui and Ng, 2000).

Several researchers have also discussed various ways in which peer review can have a positive effect on L2 writing development. Students were also reported to have increased their vocabulary repertoire when they needed to find specific words for certain expressions (Min, 2006). Moreover, peer review can help L2 student writers develop an awareness of audience (Mittan, 1989; Jacobs et al., 1998; Paulus, 1999; Rollinson, 2005) and a sense of text ownership (Villamil and de Guerrero, 1996; Tsui and Ng, 2000), contributing to their move away from writer-based to reader-based writing (Stanley, 1992).

Research has been conducted to examine the effects of peer assessment on revision and the quality of L2 writing. Several of these studies have found that L2 students can provide useful peer feedback that deals with language, content and organization (Nelson and Murphy, 1993; Villamil and Guerrero, 1998; Tsui and Ng, 2000; Hu, 2005; Min, 2005). A number of studies also found students' incorporated peer feedback into their revision and to improve their writing quality (Villamil and Guerrero, 1998; Min, 2005; Rollinson, 2005).

2.2.4 Difficulties of Peer Assessment in L2 Writing

Positive results have demonstrated the advantages of peer assessment and peer review application in L2 classroom writing practice. There have been, however, contrasting results.

Connor and Asenavage (1994), for example, found little impact of peer feedback on the subsequent revisions of drafts; only around 5% of the revisions made by their ESL writers actually resulted from peer comments.

In a survey study of 100 ESL college students' preferences for error correction, Leki (1991) found that students' peers were rated as the least helpful source of assistance with their written work.

In another survey study involving 81 ESL university students in the USA, Zhang (1995) found that, when they were asked to choose between teacher feedback and peer/self-directed feedback, an overwhelming majority of the students (93.8%) preferred teacher feedback. Zhang's findings were echoed by Nelson and Carson (1998).

In another study, Sengupta (1998) found that student writers tended to trust a teacher's feedback more than they did feedback from peers. Similar tendencies were also noted by Liu (1997, cited in Hyland and Hyland, 2006a), Nelson and Carson (1998), Tsui and Ng (2000), and Yang et al. (2006).

Several explanations have been advanced to account for these variable results, the most common of which are as follows.

Firstly, L2 students have limited knowledge of the language and its rhetorical conventions. Students have difficulties in critiquing L2 texts written by their peers because of this limited knowledge (Villamil and Guerrero, 1996). For the same reason,

L2 student writers may also lack the knowledge and skills to differentiate between valid and invalid peer feedback (Leki, 1990b; Stanley, 1992; Tsui and Ng, 2000).

One of the commonest problems with peer feedback is that students may produce 'rubber stamp advice', such as 'Be more specific!' and 'Make your ideas more smoothly connected!' (Stanley, 1992; Lockhart and Ng, 1993; Chou, 1998; Tsui and Ng, 2000). Students had difficulty in using such non-specific advice to revise their work (Liu and Sadler, 2003). As such, peer feedback was usually found to be lacking in validity and reliability (Topping, 2003).

This concern, however, is only one of the threats to peer assessment. Another major concern was the reasons learners would or would not adopt peer feedback (Falchikov, 2004). In Min's (2003) work, she found that its low adoption by EFL Chinese learners was due to misunderstanding of the writer's intentions, offering vague comments and providing no solid explanations, so that good suggestions were likely to be ignored. Based on this result, she devised systematic steps and training for the production of peer feedback. As such, she concluded that this would result in trained students providing relevant and specific comments, leading the students to accept and act upon a large proportion of the feedback (Min, 2005; 2006).

The second reason is the lack of peer assessment experience. Students' experience of assessment usually came from the teacher's summative assessment. In a traditional EFL assessment, the teacher would usually focus on the syntax of writing (Mangelsdorf and Schlumberger, 1992). In other words, from the language learning point of view, the teacher as an assessor tends to focus on the linguistic form of writing. As a result, students tend to base their assessment practice on their teacher's example. So, L2 students tend to neglect macro-textual issues and focus on surface language features in their peer reviewing (Leki, 1990b; Nelson and Murphy, 1992; 1993; Min, 2005; Al-Hazmi and Scholfield, 2007; Hu and Lam, 2010). Even when they provide comments on larger issues concerning content, organization and idea development, there is a tendency among students to provide vague comments (Liu and Sadler, 2003).

In addition, the fossilization of interlanguage, particularly for adult learners with the same level of ability, is not always an advantage when undertaking peer assessment for the purpose of language development (Long and Porter, 1985).

To pre-empt students' inclination to focus on surface issues, many researchers such as, Nelson and Murphy (1992), McGroarty and Zhu (1997), Paulus (1999) have suggested that students should be asked to concentrate on rhetorical aspects and ignore language use in formative peer assessment.

For L2 learners, however, Hu (2005) and Min (2005) had a different point of view. They encouraged students to pay attention to both micro and macro aspects of writing in their peer assessment and argued for the value of attending to the micro aspects of writing as an advantage to learning language from each other in peer assessment activity. Hu's decision was based on the understanding of her students' needs. She argues that learning the L2 is the essential task for the student. It will be difficult to ignore this basic need of students during peer assessment sessions. Min (2005) also recognised that the tendency to ignore the micro aspects of writing is problematic. She suggested that the "deliberate overlooking of grammatical errors and word usage might send an unwanted message that correct usage of grammar and words is not important as long as their ideas are fine." The message may lead the students to misunderstand the purpose of their L2 writing learning.

The third reason is students' negative attitudes. Nelson and Murphy (1992) found that some peer readers are hostile towards or over-critical of other students' writing. Besides these readers' problems, writer peers have been observed and some problems with their attitudes have been noticed. Some student writers fear being criticized for their language errors (Nelson and Carson, 1998). Guerrero and Villamil (1994) found that some students may react negatively to critical comments and become over-defensive (Amores 1997).

These attitudinal problems can create a sense of discomfort and nervousness between participants (Liu and Sadler, 2003). If such those problems exist, peer assessments could become an unconstructive activity.

The students' rhetorical tradition is another contributing factor towards their stance during peer assessment. However, the effect of this factor is still not fully understood. Some researchers argued that students from collectivist cultures (such as Chinese students) tended to avoid giving critical comments in order to maintain interpersonal harmony (Connor and Asenavage, 1994; Nelson and Carson, 1998). However, this argument has been contested by Kamimura (2006) and Hu and Lam (2010). Kamimura

argued that harmony as a non-Western cultural trait did not hamper peer feedback sessions but rather promoted it. Hu and Lam (2010) offered an explanation for the discrepancy which is based on the different ways in which peer assessment is implemented in different studies, such as responding to each other's texts in writing versus face-to-face response. It was demonstrated that the rhetorical tradition's influence is relatively insignificant compared with differences in implementation.

In the present author's experience, the rhetorical tradition is not a major factor in a student's non-collaborative stance. Instead, students' language proficiency is always the major barrier. The majority of students the author has taught in the past were facing the language barrier as well as being less keen on taking collaborative stances. This experience is supported by the findings of Mangelsdorf and Schlumberger (1992) and Min (2008). The reason, as Mangelsdorf and Schlumberger (1992) concluded, was that a collaborative stance needs more linguistic and rhetorical skill than a prescriptive or interpretive stance. This implies that peer assessment can present more of a challenge to low achievers, but this does not necessarily mean that peer assessment is not beneficial to them (Nelson and Murphy, 1993).

2.2.5 Influence of Students' Psychological Characteristics on Peer Assessment Performance

Compared to the abundance of literature with other types of experimental results, studies which experimentally examine students' psychological characteristics such as perceptions of and attitudes towards peer assessment are relatively rare.

In Gielen *et. al.*, (2010) and Hu and Lam, (2010), the student's opinions were taken into consideration. These studies suggest that students tend to have a positive attitude towards the use of peer assessment. However, the perceptions towards specific aspects of peer assessment were not reported. There was also no link shown between a student's psychological characteristics and their peer assessment performance.

To understand the impact of the student's psychological characteristics on peer assessment performance, the theory of planned behaviour (Ajzen, 1991) is adopted in the present study. According to this theory, students' behaviour is the result of three belief factors: attitudinal beliefs, normative beliefs and control beliefs. These beliefs derive from attitudes, subjective norms (the influence of significant others) and perceived behavioural control that leads to intention and behaviour (see Figure 2-1).

This theory is used to establish the link between psychological characteristics (such as perceptions and attitudes) and behaviour (peer assessment performance).

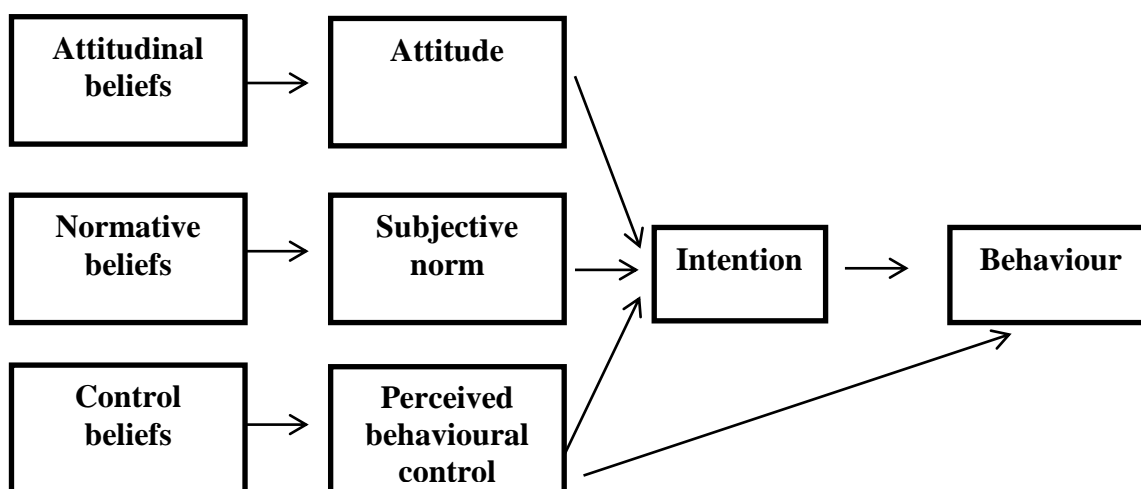


Figure 2-1 Theory of planned behaviour

2.2.5.1 Attitudinal Beliefs

One type of attitudinal belief considered in previous studies is the student's preferences concerning assessments. Some of the research mentioned below compared the students' preference for teacher assessment or peer assessment before and after peer assessment, which is effectively used to compare their attitudes towards teacher and peer assessment. Some previous studies have recognized the potential impact of assessment preferences (e.g. Zhang 1995; Jacobs *et al.*, 1998; 1999; Sluijsmans *et al.*, 2002; Hu and Lam, 2010). However, these preferences were not always given enough attention in those studies. Sluijsmans *et al.* (2002) found that the students in the experimental group (peer assessment) were more positive towards their tasks and felt more involved in the assessment than those in the control group who were not involved in peer assessment. The overall perceptions of the assessment became more positive from pre-test to post-test.

Some researchers have found that students who prefer teacher feedback have a tendency to resist peer assessment (see the debate between Zhang, 1995; 1999 and Jacobs *et al.*, 1998). Most research findings suggest that students prefer the teacher's assessment both

before and after peer assessment, even if the students have demonstrated good acceptance of peer assessment as a pedagogical method.

Going further than previous researchers, Hu and Lam (2010) tried to find a correlation between student attitudes and peer assessment performance. However, in terms of the quality and take-up rate of peer feedback, no difference was found between students who preferred only teacher assessment and those who wanted both teacher and peer assessment.

2.2.5.2 Subjective Norms

The influence of subjective norms on students in peer assessment was highlighted by Topping (1998) as well as Falchikov and Goldfinch (2000). ‘Trust’, for example, has the potential to be a key factor in students’ performance of peer assessment. The trust which students have in their own and their peers’ abilities as assessors was measured in studies by McDowell (1995), Lin *et al.* (2002), and Hu and Lam (2010). Sluijsmans *et al.* (2002) measured perceived trust in the ‘self as assessor’ and their own assessment skills, but no hard evidence was provided in relation to students’ performance. Notably, in Sluijsmans *et al.*’s (2002) study, students still did not consider they had gained sufficient skills to evaluate their peers’ work. In fact, in most of the literature (e.g. McDowell, 1995; Lin *et al.*, 2002), students expressed concerns about their own and their peers’ ability to provide constructive feedback and mark fairly.

Hu and Lam (2010) investigated the influence of trust in both self and other on learning. They found that the students who claimed that their peers’ L2 proficiency would affect their responses were not statistically significantly different from those who responded differently. In both studies by Sluijsmans *et al.* (2002) and Hu and Lam (2010), the comparative results were insignificant, although it was recognized that there were deficiencies in the research design such as small sample sizes (Hu and Lam, 2010) and inappropriate measurement methods (Sluijsmans *et al.*, 2002). The author learning from these studies, the present author has improved the research design of this study by preparing larger sample sizes and improved measurement methods (see Chapter 3).

2.2.5.3 Perceived Control

One of the obvious factors in this category studied in the literature is previous experience of peer assessment. Mendonça and Johnson (1994) identified this factor

from its effects on students' performance in peer collaboration. Connor and Asenavage (1994) also had the same concerns when they found that students who were most receptive to peer assessment had had experience with process-oriented collaborative writing.

2.3 Review of Training for Peer Assessment

Most researchers believe that there is a best practice for peer assessment training. The ideal training should be able to develop students' ability to produce effective peer feedback as well as encouraging students to collaborate.

2.3.1 Definition of 'Effective' Peer Feedback

In all of this research, the definition of 'effective' must be clarified.

Sadler's (1989:142) definition of effective feedback consists of three key elements:

[Feedback] requires knowledge of the standard or goal, skills in making multicriterion comparisons, and the development of ways and means for reducing the discrepancy between what is produced and what is aimed for.

This definition has been widely adopted by most researchers. However, varying interpretations of this definition have been applied, which is understandable since there are wide variations in the contexts in which peer assessment is applied. For example, Min (2005) had a 'step' system which aimed to develop students' ability to achieve each step of peer feedback. The definition of 'effectiveness' for each of these steps is shown in Table 2-1.

Table 2-1 Definitions of Min's (2005) four steps

Step	Definition
Clarifying the writer's intention	Students try to get a further explanation of what writers have said or what is not clear to them in the essays (e.g., an unknown term, an idea)
Identifying the problem	Peers identify a problematic word, phrase, sentence or cohesive gap

Explaining the nature of the problem	Students explain why they think a given term, idea, or organization is unclear or problematic, and which should or should not be used in the essay
Making specific suggestions	Student suggests ways to change the words, content and organization of essays.

In the above table, each step has its own effectiveness criteria. Min (2005) assumed that: if a student could carry out all four steps, he or she must have developed the evaluative skills to appreciate the standard of work. He or she should also understand the gap between the objective standards and their peer's work so that they can develop tactics to provide suggestions for modification and improvement. If a student missed one or more steps, then the effectiveness of the peer feedback would be compromised. Hence, Min's (2005) ideal for effective peer feedback was the perfection of the four steps in their peer feedback comments, based on which effective revisions could be made, as discussed further below.

2.3.2 Training Methods for Effective Peer Assessment

This section reviews the training methods reported in the most relevant studies. Aiming to improve the quality of peer feedback, Min (2005) designed a training course for a group of Chinese EFL students in the peer assessment of writing.

In this writing class, 18 Chinese EFL sophomores, who were English major students of intermediate proficiency, students were required to search for information and to reference it to substantiate their opinions in their expository essays. The students were allowed two hours to provide written comments on two of their classmates' compositions in class with the aid of a guidance sheet. Between the first and second compositions, a two-phase training programme was provided. Results were collected and compared to determine the effects of the training.

During the training, the intervention sequence as follows.

1. Writing teacher's talk 1 (4 hours);
2. Peer assessment (2 hours);
3. Teacher's talk 2 (1 hour);
4. Revision and submission.

The first phase lasted 2 hours per week for a total of 4 hours while the second phase consisted of two 30-minute teacher–student conferences outside of class.

In the first phase, a step by step model of constructive feedback was demonstrated, and in the second phase, a teacher-student conference was conducted to discuss the problematic comments raised in the first phase. Students were also reminded to follow the questions on the guidance sheets.

Min (2005, 2006, 2008) found that the training significantly increased the number of steps used in the feedback as defined in her four steps scheme. The increased usage of the steps brought improved acceptance of the comments. Min claims that the training programme was largely effective based on the fact that both the quality of comments and the quality of revisions was improved to a statistically significant extent by the training.

Min (2005, 2008) observed that more steps were used, and so more ‘specific’ feedback was produced. The feedback was claimed to be constructive and relevant in terms of aspects of rhetoric after the training. Hence, Min claimed that the training also improved the quality of the feedback. She also found that student reviewers’ stance became less prescriptive (17% decrease) and more probing, collaborative feedback increased by 12%. These results suggest that students’ perceptions had changed as a result of the training.

Hu (2005) conducted a three-year action research study of an English course, provided to upper-intermediate ESL Chinese students where classes were aimed at developing the students’ academic writing competence. The students and teacher met for two hours, three times a week, giving a total instructional time of 150 hours for 25 weeks.

Throughout the course, the students were required to complete six 500-word writing assignments, each involving a different genre. The writing and revision cycle for each assignment lasted four weeks. Each cycle contained the following activities:

- First draft
- Formative peer assessment
- Second draft
- Teacher feedback

- Third draft
- Publication

There was no clear description of the intervention in Hu's (2005) paper, due to the fact that the research contained many different sessions over three years. She found that her students showed resistance to working with their peers in the class taught in 2001. She suggested that it was necessary for the teacher to continuously support and respond to students' work throughout the peer assessment process. She designed many activities with different aims and functions which focused on aspects such as awareness-raising, demonstration, practice, reflection and instruction, explanation of procedures and pre-response review. Students were encouraged to talk about the benefits and problems of peer assessment and encouraged to reflect on the reader's stance. They were guided to discuss appropriate types of response to include in the peer review such as praising, questioning, eliciting, requesting information, pointing, clarifying, suggesting alternatives, justifying and elaborating. Students were trained under the 'scaffolding' to discuss the appropriate language for effective communication where problematic comments were shown and identified.

In her findings, Hu (2005) observed that students actively discussed each other's writing and stayed on task during the oral response sessions. Students produced quality feedback which she considered to be critical and worth taking into account for revision. The students also paid attention to the rhetorical issues and the language used in their responses. Some of the comments produced by the students even covered problems that the teacher failed to notice. Hu's (2005) work provided a wide range of ideas and activities for peer feedback training. In a similar way to Min (2005), she also introduced useful criteria and helped students to internalize them through teacher-student conferences.

Compared to those in Min (2005), Hu's (2005) ESL students were at a higher level of proficiency and were required produce a 500-word essay as opposed to 260 words for Min's students. Hu also claimed that she had positive results for the quality of peer feedback received as a result of her training programmes. Changes in students' collaborative stance were also mentioned and changes in their attitudes towards collaboration were also noted.

Compared with Hu's work, the author of the present thesis is more interested in Min's (2005) training methods for two reasons. First, it is a detailed system which was developed based on previous work including that of Stanley (1992), and Zhu (1995). Secondly, Min specified more details of her training methods than other researchers, so that it could easily be referenced during course design.

2.4 Discourse and Interaction

The previous sections have reviewed the literature on peer assessment research. Under the influence of socio-cultural learning theories, the present author recognizes that pedagogical discourses should focus on constructing an effective socio-cultural interaction model of peer assessment. Hence, the interest of successful peer assessment lies in the discourses and interactions.

In 1996, Bernstein published *'Pedagogy, symbolic control and identity.'* This book established the model for pedagogic communication, and proposed several important concepts which are useful for analyzing the discourse and interaction in peer assessment.

2.4.1 *Regulative Discourse and Instructional Discourse*

Bernstein (1996) considered that pedagogic discourse refers not only to the subject's content and competence in it, but also to the transmission and evaluation of knowledge; that is, it refers to *the what* that is transmitted, *how* it is transmitted, and also how the subject should be received or understood. The essential idea within Bernstein's theory is that pedagogic discourse is made up of two components: regulative discourse and instructional discourse. Regulative discourse delivers the dominant values of society and regulates the form of the knowledge transmitted whereas instructional discourse delivers the competence of the content transmitted. Due to the nature of these two types of discourse, the regulative discourse always dominates the instructional discourse.

2.4.2 *Classification and Framing*

Another important pair of concepts is classification and framing. According to Bernstein (1996), these define the boundaries of the categories within a discourse. Hence, they define the structure of the discourse.

Classification refers to the boundary strength between what is classified (Bernstein, 1971). Strong classification means strong boundaries, such as clearly separated subjects, such as English, Maths, and Physics, while weak classification refers to weak boundaries between the subjects. Strong classification is characteristic of the collection code; and weak classification is characteristic of integrated code.

Framing refers to the message system of pedagogy (Bernstein, 1971). Do teachers and pupils control content, how it is organised, how it is sequenced, and so on? A syllabus with rigid topics, to be completed in a predetermined order, within a specified time, would represent strong framing. Weak framing would occur when the teacher is able to select topics on the basis of pupil interest, or some other such principle, and students may provide input in organising the sequence and pacing of material according to their readiness and stage of development.

Corresponding to Bernstein's (1996) definition of discourses, there are two types of framing: instructive and regulative framing. Instructive framing communicates the rules, the content, and the selection of knowledge which is to be infused as the teaching goal in the interaction. For example, the 4-hour introduction during Min's (2005) training for peer assessment can be considered as the effort to build up to the instructive framing, during which detailed instruction of how to use the 'steps' in constructing peer feedback was introduced. Regulative framing communicates the participants' position and responsibility, and worked as the interactive rules sets in the interaction. For example, the teacher-student discussion held by Hu (2005) can be regarded as a method to establish regulative framing for peer assessment. During this discussion, the students familiarise themselves with the roles they will play in peer assessment.

In this work, the strength of the framing is judged by the teacher's evaluative actions during the interaction. A clear evaluation given by the teacher is counted as the strongest framing, and a non-evaluative comment is counted as the weakest. Details of the strength of framing are given in Table 3-7 in section 3.7.5. The notions of framing strength, such as IF++ and RF++ are also explained in section 3.7.5.

2.4.3 Convergent and Divergent Assessment

Based on the concept of the strength of classification and framing, Pryor and Crossouard (2008) established the concept of convergent and divergent assessment. In convergent assessment, the classification and framing are both strong. The teacher aims

to find out whether or not the students have the required knowledge of pre-determined subjects and the conversation usually follows the initiation-response-feedback/evaluation pattern. In divergent assessment, the classification and framing can become weak. The teacher aims to find out what students know about the required knowledge of pre-determined subjects and open-ended questions and exploratory responses usually occur in the interaction.

Rather than seeing these two conditions as a fixed practice situation, Pryor and Crossouard (2008) argued that the strength of classification and framing could oscillate on this convergent-divergent continuum. Depending on the teacher or students' needs, teachers can choose to move flexibly and fluidly along the continuum, employing four different teacher identities; namely, teacher, assessor, subject expert and learner.

According to Pryor and Crossouard (2008), the effectiveness of formative peer assessment is determined by changes in the teacher's classification and framing. Such changes should be along the convergent-divergent continuum. This latter theory can serve as a useful tool in analysing previous studies. Min (2005) found that students' language improved in their writing and feedback, which demonstrated the effectiveness of the protocol provided in the teaching. This result could be recognized as having the characteristics of convergent assessment. In Hu (2005), the teacher established the learners' identity during a discussion in which she reflected that "they [students] supplemented her [the teacher's] comments on many occasions or anticipated many of her [the teacher's] suggestions, thus demonstrating the strong development of learner's understanding [of the context] and helping her [the teacher] develop her [the teacher's] own in turn." This suggests that divergent assessment took place where the teacher took the role of learner and encouraged students' feedback.

The two examples above demonstrate the typical scenarios in reality as described in Pryor and Crossouard's (2008) theoretical system. As in Pryor and Crossouard's (2008) claim, the learning should occur during the process of switching (oscillating) from the firm shell scenario (convergent) to the soft core scenario (divergent). In the meantime, the teacher's role should be flexible during this process, also switching from the normal teacher's role to a learner's role.

The ideal scenario would be difficult to create. In Pryor and Crossouard (2010), there was further discussion of the teacher's roles during convergent and divergent

assessment. The convergent end of assessment began with the teacher's role as teacher and assessor, then taking on the role of teacher and knowledge expert, and finally that of learner where the assessment reaches the divergent end. During this process of role change, it was claimed that students were engaging in four different levels of learning:

- Completing the task in hand;
- Making sense of criteria;
- Thinking about improvement;
- Invoking learner identities for peer assessment;

Each level was recognized as superior to the preceding level. To complete the task in hand and make sense of the criteria, both teacher and students produced feedback focused on concrete/procedural and practical aspects. To think about improvement, feedback from both the teacher and students should encourage reflection in relation to the students' work to their discipline and social rules.

Invoking learner identities in formative assessment refers to the highest level of the student's understanding of the peer assessment. The student would be able to understand the task in peer assessment spontaneously, switching their roles among learner, assessor or subject expert in the most appropriate manner. Most importantly, the student will be able to permanently establish the identity of a peer assessor, so that the student could be highly autonomous during peer assessment.

The steps designed by Min(2005) corresponded closely to the above levels. She recommended that if the students could complete the four steps in their peer feedback comments, then the peer assessment training could be regarded as successful.

Based on that assumption, Min's (2005) students were reported as seeing themselves as reviewers who had 'changed the way' they delivered the content of their feedback. Students were able to self-monitor their own problems and had confidence in reviewing others' work. The students were also reported as seeing themselves as a writer, and acknowledged the value of learning from their peers concerning how to focus their ideas and view things from different perspectives.

In Hu's work (2005), although there was no agreement like the Pryor and Crossouard's (2008) four levels, she reported that her training had led to agreement. The students showed their understanding of writing as a recursive process. They also realized the importance of revision, and adopted a more critical attitude towards their own writing.

Hence, Pryor and Crossouard's (2008) model is partially matched by Min's (2005) and Hu's (2005) earlier work.

However, the match was only on the surface level. Neither Min's (2005) nor Hu's (2005) explicitly demonstrated the change between convergent and divergent assessment. On the contrary, Min's works mainly landed on the convergent end, but Hu's (2005) work was more closely linked with divergence. Min's (2005) and Hu's (2005) work was heavily focused on technical aspects, with their efforts directed at creating a solution which could potentially become the standard practice when introducing to peer assessment in EFL/ESL writing. Hu's (2005) work was mostly empirical; multiple activities were designed and tested for their effectiveness. Both authors referenced socio-cultural theories, but neither of them developed a theoretical framework for their models.

2.5 Teacher's Role in Classroom Peer Assessment

According to Bernstein's (1996) theory together with the further development by Pryor and Crossouard (2008), successful formative peer assessment should be achieved under the following conditions:

- Suitable framing and classification settings;
- Flexibility in teacher roles

So that:

- The peer assessment tasks have clear boundaries;
- The students can access sufficient information;
- The students are well motivated;
- The students are placed in a power relationship conducive to providing peer feedback.

One of the key aspects of interaction between the teacher and students is the roles they adopt. According to Pryor and Crossouard (2008), it is essential for the teacher to adopt the appropriate role in order to engage with the students in the teaching discourse. The roles adopted by the teacher and students are related to the types and strength of framing in the classroom interaction. These roles are considered differently in the literature.

From the socio-cultural perspective, typical roles in the interaction can be represented by: teacher as authority (strongest framing), subject expert (intermediate strength) and learner (weakest framing). The students also undergo an appropriate role transformation

during the interaction. In the ideal approach, the teachers' role should be flexible for the different purposes of formative assessment to opportunistically provoke different levels of responses from the students. However, such an ideal is difficult to achieve. This was one of the reasons for the technical approaches adopted by Min (2005) and Hu (2005).

2.6 Research Gaps

A detailed review of the relevant literature has led to the identification of the following gaps in knowledge:

1. Although a number of pioneering research studies have proposed various methods which are claimed to be effective in training for peer assessment, no theoretical underpinning was provided for the proposed methods.
2. Previous research has relied heavily on quantitative results as evidence of the effectiveness of training methods. Little qualitative analysis was carried out on actual classroom interactions.
3. Student's beliefs and opinions have been considered in previous work. However, the links between attitudes and perceptions and behaviour patterns and peer assessment performances were not clearly established.

The current study is intended to address those gaps.

2.7 Summary

Previous studies, both experimental and theoretical, were primarily based on Bernstein's theoretical formulations, which derive from Vygotsky's psychological theories of learning. Recent developments in this context were mostly based on empirical (e.g. Hu, 2005) and experimental (e.g. Min, 2005) results. The theoretical developments were in two directions: one was presented by Topping (2009) and Flachikov (2001) whose approach was based on social-psychological theory; and the other by Pryor and Crossouard (2010) with their approach based on socio-cultural learning theory. In the present work, the author mainly makes reference to the socio-cultural approach towards peer assessment. Based on Bernstein's theory, classroom interaction is examined in terms of its classification and framing in order to investigate the underlying mechanism of the context.

The formulation by Pryor and Crossouard (2008) of convergent and divergent assessment is also considered. The research examines the process of learning by identifying the nature of the assessment. The framework of the learner as becoming a peer reviewer is shown in figure 2-2. This framework includes 2 teaching approaches, 2 teaching discourses, and two stages of student results and performance. The purpose of these assessment discourses in the left hand column corresponds to the expected results in terms of student performance in the right hand column. At the end of the training, students' skills in producing peer feedback will be reflected in their opinions, attitudes and perceptions towards peer feedback activity and the step training. Moreover, additional consideration of the psychological aspects of student performance will be taken into account. The theory of planned behaviour is used to provide an additional perspective concerning the experimental outcome.

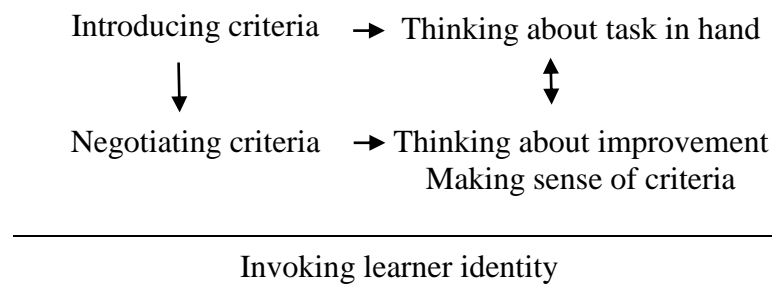


Figure 2-2 Conceptual framework

Chapter 3 Methodology

3.1 Aims

The first two chapters have introduced the background to the research and reviewed the relevant literature. After reviewing previous studies in the field, a gap between theoretical ideals and current practice has been identified. This work aims to fill this gap by conducting research into training for formative peer assessment in L2 writing classes.

For clarity, the research questions are repeated below:

1. What is the nature of peer feedback provided in both groups?
 - a. To what extent does peer feedback follow Min's four steps as a result of Min's training method?
 - b. What is the quality of students' argumentative writing?
 - c. What are the students' attitudes towards and perceptions of peer assessment before and after the training?
2. How does the teacher model and negotiate to shape students' peer feedback?
 - a. Based on Bernstein (1996), how explicit is the modelling and negotiation in providing guidance to the students?
 - b. How does the teacher engage students in the discourse of modelling and negotiation?

To answer the research questions the following decisions must be made.

What should be evaluated to measure the effectiveness of peer assessment training?

In the literature, different methods of peer assessment training are claimed to be effective. These claims are further supported by quantitative evidence. However, when the present researcher tried to reproduce these training results in the classroom, observations and feedback gave rise to concerns regarding the claims for the effectiveness of these training methods.

Despite the difficulty of reproducing the results for effective training, previous studies also provide little theoretical explanation for the effectiveness of training. For this reason, it is difficult to predict the outcome of a similar method applied to a different classroom setting. For the same reason, it is difficult for a teacher to use existing

theories to develop their own peer assessment training based on the reported successful examples.

In order to fill this gap, it is proposed that both quantitative and qualitative analyses should be performed to provide further understanding of the effectiveness of training. The measurement of effectiveness should not only focus on statistical evidence and empirical observations, but also on the individual student's opinions, progress and involvement. The effectiveness of a training method should be able to be explained by theory, so that teachers and researchers could further develop their training based on the proposed successful example.

What are the perceptions and attitudes of students towards the peer assessment and training methods?

This study investigates student's perceptions of and attitudes towards several aspects of the peer assessment practice. The data were collected using questionnaires and interviews.

A link between the measurements of student's perceptions and attitudes and their performance in peer assessment can be established by using the theory of planned behaviour (TPB). These perceptions and attitudes will not only become indicators of the training and the peer assessment activity's overall effectiveness, but can also be used as reference for future peer assessment training design and improvement. The interview results are qualitative indicators and evidence of students' perceptions and attitudes towards the training and peer assessment practices. It is expected that the information provided by the interview will contain detailed information which the questionnaire results cannot reveal.

What are the effects of socio-cultural factors in the discourses used?

Based on Bernstein's (1996) concepts of classification and framing, qualitative measures were applied to identify the important socio-cultural factors in the discourse and these socio-cultural factors are matched with the observations of the EFL writing classroom teaching. Moreover, the effects of the 'alternation' of 'framing strength' as suggested by Pryor and Crossouard (2008) in the classroom interaction are examined.

These findings will provide detailed evidences of how modeling and negotiation discourse was realised through interaction. More importantly, the analysis will

demonstrate the effects of socio-cultural factors on the interaction, and will show what kind of interaction is effective in terms of engaging with students and the reasons for such effectiveness. This will help in answering the second research question.

3.2 Critique of Methods Used in Previous Work

In the literature, different research methods were used. The following section reviews some representative examples relevant to this thesis.

Research methods in recent studies on peer assessment training methods are primarily quantitative, aimed at examining the validity and reliability of the test or assessment. Recently, the concept has been revised as in response to the positivist epistemology and the assessment literature (language testing in particular) (Lumley et al, 2005). Qualitative methods are employed to look beyond the validity of test scores in terms of “the validity of inferences made about test-takers on the basis of their scores, that is, the meaning of scores” (Lumley et al, 2005).

Dunn and Mulvenon (2009) reviewed nine research studies to evaluate their use of statistical methods. In those studies, statistical methods were employed to develop a predictive model. Dunn and Mulvenon (2009) cast doubt on this methodology to support conclusions concerning the ‘effects of formative/peer assessment in learning achievement’. They criticized the fact that sound evaluation practices and statistical methodology seemed to be underdeveloped, as the focus was on “completing a *prediction model* with a set of data, or the generation of frequency tables was represented as analyzing student achievement”, rather than on “the type of assessment, its design, and psychometric properties” (Dunn and Mulvenon, 2009: 9). The same criticisms can be made of preceding research, namely that of Min (2005, 2008, 2010), Hu (2005) and Zhu (1998).

Qualitative methods such as discourse analysis and ethnographic methods are also commonly applied to investigate assessment practice and its social consequences.

Discourse analysis, in this context, is used to understand how formative assessments of this kind help students to learn. For example, Leung and Mohan (2004) developed a simple decision-marking analytical model which helps to understand how elements of assessment for learning are co-produced by teachers and students. Discourse analysis can reveal subtle details of an interaction through analysing the elements contained in

the interaction. Its advantage is that the analysis can be based on various sources such as dialogue notes, excerpts, sound recordings and video clips. This helps in the investigation of teaching and learning from a social-cultural perspective which is expected to suggest fundamental development and improvement to the training.

In recent years, there is also a trend towards mixed methods, where both qualitative and quantitative data are analysed. Quasi-experimental research by Zhu (1995) adopted a mixed-method approach to analyse the effects of peer assessment training. Evidence of the quality of students' writing was collected, their ability to critique the writing of peers, and their attitudes towards peer revision. By using mixed methods, Zhu performed both qualitative and quantitative analysis of the peer assessment training. To investigate students' ability to critique, for example, quantitative analysis was used to calculate the frequency of critical comments, while qualitative analysis was used to understand the experiences of instructors and students in the process of producing critiques. Min (2005) also used a mixed-methods approach in her case study. She collected students' peer feedback, written work, and learning journals and conducted interviews and in her analysis of the effects of training triangulated evidence from quantified students' feedback work, and data on students' attitudes toward peer revision and writing from a learning journal. In both Zhu's (1995) and Min's (2005) work, their findings were derived from different sets of data but the methods complemented each other making it possible to clarify and illuminate their research findings. Although Min (2005) did use a mixed method, she relied heavily on quantitative analysis to prove the effectiveness of the training.

It can be concluded that, in previous studies on peer assessment training, mixed methods have only been used for the evaluation of the effects of training in a specific context. Therefore, the use of qualitative methods and analysis is still very traditional and lacks a social-cultural perspective. Qualitative data collection and analysis using a social-cultural perspective can provide a profound understanding of the teaching and learning, the advantage and disadvantage of qualitative methods and analysis should be considered for adoption to reveal the fact about the impacts of social cultural factors on teacher's interactional management and students learning progress.

3.3 Methodological Basis

3.3.1 *Rationale for Research Methods*

Qualitative and quantitative paradigms represent contrasting views in terms of epistemology, ontology and axiology (Johnson and Onwuegbuzie, 2004). The quantitative paradigm, or positivism, advocates retaining objectivity in research. This means that time- and context-free generalizations (Nagel, 1986) are desirable and possible and the true causes of social outcomes can be determined.

The qualitative paradigm refutes the positivist viewpoint and argues that it is impossible to remain absolutely context and bias free. Working under different ontological, epistemological and axiological assumptions about the goals and nature of research advocates of, both schools of thoughts assert the incompatibility thesis (Howe, 1988), which is that qualitative and quantitative paradigms, including their associated methods, cannot and should not be mixed.

However, in order to gain a comprehensive understanding of educational and social phenomena, Johnson and Onwuegbuzie (2004) argue that taking a pragmatic and balanced approach to the use of methods would generate more fruitful results with which to advance knowledge.

Pragmatism focuses on practical problems and considers that truth is what works for a specific problem at a particular time. The action taken and its consequences constitute the meaningful conception of the subject studied and therefore provide understanding of the studied subject (Barbalet, 2009). Since pragmatics deals with practical problems rather than issues of reality and theories of society such as in the traditional paradigms, a middle ground is advocated for the use of mixed methods. By doing so, the inherent bias in methods originating from a traditional paradigm can be avoided (Johnson and Onwuegbuzie, 2004) which helps to improve the validity of the research (Bryman, 2007). While quantitative methods aim to achieve statistically reliable and generalizable results, qualitative methods can be used to explore the details of interactions and provide a rich account of perspectives. In addition, when triangulating data, agreement in the results from different methods can yield greater insights (Greene, et al., 1989). In this study, a pragmatic stance is taken to explore the adoption of a particular training method for peer assessment.

3.3.2 Nature of Research

Having explained the rationale for taking a pragmatic view and adopting mixed methods for this research, it is important to elaborate on the nature of this research study, which can be categorized as evaluative research. This type of research involves the evaluation of an intervention in an organizational program with the aim of determining whether it has achieved its anticipated goals (Bryman, 2008). A typical approach in evaluation research is quasi-experimental, due to the impracticality of a strictly controlled experimental design with random sampling, and the ethical issues which arise when assigning research participants into the different groups.

Additionally, the qualitative approach is based on a recognition of the importance of an in-depth understanding of the context in which an intervention occurs. While this study aims to improve students' abilities in constructing peer feedback through step training, it also aims to evaluate the effects of the training using both quasi-experimental and qualitative methods to answer two different research questions. While one of these questions deals with the quality of students' work from a quantitative point of view, the other question deals with the process of students' learning and the context where the learning takes place from a qualitative point of view.

As a result, through the use of mixed methods in evaluation research, a systematic understanding of the nature of the research objects and a holistic understanding of 'how' the peer assessment training facilitated students' development of peer feedback skills and their performance in the interactional context can be developed.

3.4 Research Design

The present author's classes in this study had lower levels of achievement in English than Min's (2005) class; however, the author's class can be considered to be a typical EFL class, as more frequently found in Taiwan. Conversely, Min's (2005) class was one of the four most prestigious universities out of more than 100 higher education institutions in Taiwan (the National Cheng Kung University). Also, Min's class majored in English Studies, but the two classes in the study reported in this thesis did not major in English, but in Japanese. Hence, it is believed that the classes investigated in this study are likely to be more representative of typical EFL classes.

In Min's (2006) experimental research, the work of one group of students before and after peer assessment training was analyzed to show the effectiveness of the training. However, in this thesis, two independent groups are randomly assigned to be experimental and control groups. The work produced by these groups was then analyzed and compared to show the effects of the formative peer assessment training. The advantage of this two-group design is that repeated comparison is possible between the two groups after the training tasks in different writing topics. In Min's design, this comparison could only be performed within-groups. The present research avoids this deficiency by using a control group.

3.4.1 Overview of Data Collection Procedures

The author designed various peer feedback training programmes for the two classes. Data was collected using the following methods.

- Classroom sessions were video-taped with consent given by the school and students.
- Students' peer feedback work was collected.
- Students' essays were collected.
- Transcripts from interview with the students were collected.
- Questionnaire results were collected.
- Classroom observation notes were filed.

The overall plan for data collection is detailed in the following table.

Table 3-1 Date collection procedures

	Class	Data source	Week 1	Week 5/6	Week 9/10	Week 11	Week 13
Students' written peer feedback	A	Whole class		✓	✓		
	B	Whole class		✓	✓		
Students' argumentative writing	A	Whole class		✓	✓		
	B	Whole class		✓	✓		

Questionnaires	A	Whole class	✓			✓	
	B	Whole class	✓			✓	
\Interviews	A	Six randomly chosen members		✓	✓		✓
	B	Six randomly chosen members		✓	✓		✓
Video recording	A	Whole class	Week 2-13				
	B						

3.4.2 Preparation Activities

The author discussed the prepared training programme and the quasi-experimental design with the teachers, who were lecturers of two undergraduate classes in a language college in Taiwan and the quasi-experiment with peer assessment was implemented in their English writing classes which were part of the existing syllabus. Discussions took place on several occasions before these actual experiment started. During the discussions, details of the quasi-experiment and its scope were clarified. Peer assessment activities and training sessions were planned and scheduled. It was agreed that the peer assessment training would be delivered separately from writing, so that the boundaries of the experimental training could be clearly identified.

It was also agreed that the nature of the training activities would be differentiated between the two classes. The training method in one class had to contain the ‘step’ training design, while in the other class the training had to exclude the ‘step’ training design. It was also agreed that all of the same writing instruction had to be delivered. Students’ target for writing practice set by the school had to be met. It was decided that the peer assessment activity would take place at the same time as the scheduled teachers’ assessment. The teachers also had to make amendments to their teaching plans in order to adopt the author’s training plan. The finalised teaching plan is shown in the following table.

Table 3-2 Teaching plan

Week	Session	Class A Teaching Brief	Class B Teaching Brief
1	1	Orientation	Orientation
	2	Introduction to topic 1	Introduction to topic 1
2	1	Discussion of topic 1	Holiday on school calendar

	2	Writing instruction	Holiday on school calendar
3	1	Writing instruction and brainstorming	Holiday on school calendar
	2	Teacher's feedback and writing in class (and as HW)	Holiday on school calendar
4	1	Peer assessment training: step training task 1	Discussion of topic 1
	2	Peer assessment training: step training task 1	Writing instruction
5	1	Peer assessment	Writing instruction and brainstorming
	2	Teacher's feedback and writing in class (and as HW)	Teacher's feedback and writing in class (and as HW)
6	1	Introduction to topic 2	Peer assessment training: guidance
	2	Discussion of topic 2	Peer assessment
7	1	Writing instruction	Teacher's feedback and writing in class (and as HW)
	2	Writing instruction and brainstorming	Introduction to topic 2
8	1	Teacher's feedback and writing in class (and as HW)	Discussion of topic 2
	2	Peer assessment training: step training task 2	Writing instruction
9	1	Peer assessment training: step training task 2	Writing instruction and brainstorming
	2	Peer assessment	Teacher's feedback and writing in class (and as HW)
10	1	Teacher's feedback and writing in class (and as HW)	Peer assessment training: guidance
	2	Introduction to topic 3	Peer assessment
11	1	Further discussion of topic 3	Writing in class (and as HW)
	2	Writing instruction	Introduction to topic 3
12	1	Writing instruction and brainstorming	Further discussion of topic 3
	2	Writing in class (and as HW)	Writing instruction
13	1	Peer assessment	Writing instruction and brainstorming
	2	Teacher's feedback and Writing in class (and as HW)*	Teacher's feedback and writing in class (and as HW)
*End of experiment			

The author's requirements in this study were as follows. Peer feedback would be used in the teaching as a significant part of the assessment of the learning activity, and the methods intended to improve student's peer feedback would be examined in the teaching. Teachers would take full control of the classroom teaching, including interventions and the choice of teaching materials. Meanwhile, the author would be

present as an observer and would have no influence on how the teacher would deliver the teaching plan. This agreement was designed to allow the teacher to intervene as felt appropriate without undermining their control over the direction of teaching.

A limiting factor in the experiment which the author had to agree to was the specific teaching material used. The school had already designated the main text book as *Send Me A Message* by Mackey (2006), and every student had already purchased a copy. *Practical English Writing Skills: A Handbook with Practice* by Scheraga (1999) was considered more appropriate by the author, but as this was not compulsory material few students owned it.

3.4.3 Course Design

The course that incorporated the quasi-experiment was called Practical English Writing (PEW), conducted in year 4 in the second semester. PEW met for two hours twice a week, giving a total instructional time of 52 hours for the semester. The aim of the course was to help students to write a short article in clear English.

The pedagogy adopted in PEW can be described as process-oriented, genre-centred and theme-structured. After the inclusion of the author's experimental content the major instructional activities were writing instruction, multiple-drafting, peer review and revision.

The writing instruction was designed to include activities such as reading on specific themes, mini-lectures on argumentative writing in relation to audience, purpose, rhetorical structure and support, and pre-writing tasks such as brainstorming and discussion. These activities emphasized writing as a social act and as a recursive process of "overlapping and interconnecting stages" (Nelson and Murphy, 1992: 172). They were aimed at helping the students go through the stages of composition and to acquire strategies for producing effective argumentative writing.

During PEW the students were expected to complete three written assignments of 120-150 words each on different topics. The writing cycle for each assignment lasted six weeks and consisted of three stages: draft-peer review; draft-teacher; and third draft submission. Due to the length of the course, these writing cycles overlapped in places.

Table 3-3 summarizes the major activities involved in each writing cycle along with the implementation of the tasks for the training in formative peer assessment. The peer

assessment sessions were designed to give ample time in which to produce detailed, carefully considered feedback of one hour for oral and written responses. At least four days was also given to students to revise their drafts before resubmitting, if necessary.

The inclusion of oral and written responses was intended to give the students opportunities to take advantage of the benefits that both types of response could afford (see Hedgcock and Lefkowitz, 1992; de Guerrero and Villamil, 1994; Liu and Sadler, 2003). It was also anticipated that the social interactions involved in the oral responses could generate “a wealth of activities, strategies, and behaviours” (Villamil and Guerrero, 1996: 69) that could activate and enhance the cognitive processes essential for acquiring effective discourse strategies and writing skills. Written responses, on the other hand, allowed students extra time to think through their comments and suggestions (Zhu, 1995). During the formative peer assessment, students worked in pairs, which were fixed once the pairings had been made on the basis that this would encourage interaction and dialogue. In the case of one student who could not attend, the teacher would introduce a group of three to cross-examine each other’s work.

The students were encouraged to pay more attention to the macro aspects of writing, for two reasons; the first of which was, in response to Hu and Lam’s (2010) call for research to investigate efficient ways of developing students’ ability in the macro aspects of writing, such as organization, content and format. Secondly, students were allowed to use their L1 in their feedback in case they could not spontaneously produce the appropriate expression in English. As students were allowed to use this type of comments, the benefits of language development through peer feedback were considered to be limited (see section 2.2.4 for more discussion). As a result, in considering the weight of the cognitive benefits from critical thinking over language development, students were encouraged to concentrate on the macro aspects of writing during the class time.

Table 3-3 Course design

Week	The Writing Cycle Activity	Peer Assessment Training Activity	
1/6/10/11	Introduction to the topic		
2/7/12	Start of an assignment; writing instruction		
3/7	Writing instruction		

4/8	Writing instruction	Class A: Task 1 and Task 2	
5/9/13	Exchange of 1st drafts, oral response and written comments to 1st drafts, submission of 1st and 2nd drafts together with written peer comments		
6/10	Return of drafts and teacher feedback. Submission of 3rd drafts for final marks		Class B: peer assessment guidance without steps training

3.4.4 Peer Assessment Tasks

The peer assessment activities were centred on two tasks:

Task 1: Introduction of criteria. This required students to work on peer assessment with existing criteria introduced by the teacher. This task was the basic task of the peer assessment practice. The teachers introduced the criteria for peer assessment by giving relatively detailed instructions and also clarified the scope of the work of peer assessment. During this task, the students were assisted by a worksheet, the design of which varied by class. The major differences were that in class A the course was delivered with the content of ‘step’ training, with reminders of the ‘steps’. In class B, the course excluded the ‘step’ training and there were no reminders of the steps. The worksheet contained questions and guidelines which invited the students to provide their peer feedback (see Appendix B).

Task 2: Negotiation of criteria. This task required students to work on peer assessment with the criteria agreed by negotiation with the teacher. In this task, students were further invited to provide peer feedback regarding their previous work in Task 1. The negotiation between teachers and student concerning the criteria was closely related to their experience of Task 1. However, the author and the teachers all regarded this task as superior to Task 1. Although assistance from the teacher was still available, the students had to rely on their own decisions to complete the task. One of the targets of the second task is to increase the chances of invoking learner identity for peer assessment. More details of the tasks are presented in section 4.5.

3.5 Participants and Sampling

This section clarifies the details of the quasi-experiment setting.

Students

There were 127 students involved in this study, divided into two classes.

Class A: led by teacher Eva (pseudonym), majoring in Japanese studies.

Class B: led by teacher Ann (pseudonym), majoring in Japanese studies.

Institution

The W College of Languages (pseudonym, now known as the W University of Languages) is a well-established school in southern Taiwan at which five foreign languages are taught. The majority of students register on four-year undergraduate courses after graduating from senior high school or vocational high school. On satisfactory completion of the course, the students are awarded a Bachelor's degree. Under this system, all students are required to collect 36 English course credits and fulfil the graduation benchmark of CSEPT (College Student English Proficiency Test) for a score above 260 in total (equivalent to IELTS 5.5).

Instructors

Two teachers were involved in this study.

1. Eva holds a PhD degree in English teaching, and was in her first year of a teaching contract in the graduate school of the W College of Languages. She previously had over 5 years of teaching experience in a vocational college.

She had written a PhD thesis on portfolio assessment and peer assessment, which was used in her pedagogy and so she had a good understanding and experience of applying peer assessment tasks and guidance. For her, the benefits of using peer assessment in class were the exchanges of ideas and opinions and the subsequent improvements this had on the standard of writing. From the peer worksheet she had designed for students in a previous semester, it seemed that Eva had followed what the literature suggests in guiding students on both the micro and macro aspects of writing.

2. Ann holds a masters degree in English teaching and had 16 years experiences as a teacher holding a permanent position in the W College of Languages. She was also awarded the 'Outstanding Teaching' award by the Ministry of Education in Taiwan in 2009.

Ann had less experience in the practice of peer assessment. She regarded the purpose of using peer assessment as to reduce students' errors before final submission. She did not, therefore, make a significant effort to introduce peer assessment skills to the class. Instead, the peer assessment applied by Ann in the past had focused primarily on micro corrections, in line with the teacher's requests.

Classes

In class A, the 65 students' English language proficiency test scores were on average 180.19 (standard deviation 53.15, equivalent to CBT TOEFL 173 or IELTS 4, see 'The Letter of Language Exam Equivalency Standard 0930123968A' by Ministry of Education of Taiwan). The instructor (Eva) pointed out that the students in this class were still at a 'low level' in English: "For example, most of the students would use a new word in their electronic dictionary without considering the context and collocation of the words. Their writing was difficult to understand due to the presentation of their expressions. They thought and expressed in their mother tongue and selected words from the target language accordingly. There were also spelling errors and typical mistakes such as run-on sentences. The logic between sentences was usually not clear." Eva also pointed out that the students were low in confidence about learning English despite the fact that they all understood its importance. On the other hand, these students were considered to be well disciplined and well behaved compared to her other classes, and showed motivation to improve their English proficiency.

In class B, the 62 students' language proficiency test scores were on average 182.17 (standard deviation 54.47). According to the class instructor, Ann, the students were "generally low in language proficiency and showed low motivation to improve their English". Their use of language was "poor and with typical sentence problems."

Gender Ratio

In both classes, there were more female than male students. In class A, there were 7 male and 58 female students. In class B there were 14 male and 48 female students.

Retake Students

Students who failed to meet the examination criteria of a 'pass' had to retake the module. There were 10 retake students in total: 4 in class A and 6 in class B.

3.6 Justification of Data Collection Tools

3.6.1 *Students' Written Peer Feedback*

Worksheet Design

There are three parts on the worksheet (see Appendix B).

The first part invites students to give feedback on general aspects of their reviewed works. The second part asks students to provide peer feedback using four statements followed by guiding questions. In class A, there were sample answers with a reminder of the 'steps'. In class B, there were no sample answers. In the spaces given, students were invited to clarify the writer's intention, identify problems, explain the nature of the problem, and make specific suggestions. The third part was a space for the writer's second or third revision.

Two assignments were successfully collected. There were sixty copies for topic 1 and 55 copies for topic 2 from class A. From class B, sixty-two copies were collected for topic 1 and 58 copies for topic 2.

Students could use the worksheets during their peer assessment practice. Before filling in the worksheet, the students first read and discussed each other's work. Students were given 30 to 50 minutes to discuss and complete the worksheet.

3.6.2 *Students' Argumentative Writing Assignments*

The main purpose of this course was to improve writing. The peer assessment practice was based on the writing produced by the students. So, the writing was the starting point and the final output (apart from the peer feedback) of the peer assessment. The students were asked to write two argumentative writing assignments. The topic was based on an existing claim contained in reading material and the students were expected to identify the defects in the claim and make reasoned arguments against it. The first topic was 'disagreement with a non-smoking policy in the building' and the second topic was 'disagreement with a company merger case'.

Students were expected to compose an argumentative essay of 150-200 words. The writing should summarise their point of view, the argument against the claims and the final conclusion. Each writing assignment was written as a first draft during the class hours. The students would then peer assess each other's work in pairs. After the peer

assessment, the students were asked to revise their work in response to the peer feedback. The revised work was then compared with the first draft. The comparison was expected to demonstrate improvement due to the peer assessment.

3.6.3 Questionnaire

The questionnaire was designed to measure students' perceptions and attitudes towards various issues before and after the peer assessment intervention. Based on the theory of planned behaviour, the questionnaire was intended to identify and assess the relative importance of the characteristics of the target behaviour. In the light of the discussion in section 2.2.5, this study involved the use of the questionnaire prior to and after the peer assessment intervention.

The main principles on according to the dependent variables were derived were based on previous research (Wen and Tasi, 2006). Questions were designed in accordance with different groups of beliefs and in each of these groups detailed questions were targeted at specific opinions concerning the practice of peer assessment. The main questions were designed as multiple choices with answers as grade marks from 1 to 5.

Two sets of questionnaires were used before and after the peer assessment. In this thesis, they are denoted as the 'pre-training' and 'post-training' questionnaires respectively, or pre or post questionnaire for short. Most of the questions were identical in both questionnaires. However, in the post-training questionnaire, some questions about the peer feedback activity were added. The questionnaires can be found in Appendix E.

The first set of pre-training questionnaires were given to both classes of students at the beginning of the course, and the set of post-training questionnaires were distributed 10 weeks later when the course was about to finish. Both questionnaires asked for students' names and other personal details but added that giving this information was optional. The results were collected and recorded using a fully symbolized system where the students' personal details were coded with numbers or symbols in order to ensure that privacy would not be compromised in the research.

3.6.4 Interviews

Interviews were used to investigate students' and teachers' detailed opinions. This method was used to explore the reasons for perceptions, attitudes and behaviour from a qualitative point of view.

The students were either randomly selected for interview after the classes or invited to attend the feedback sessions of the course on its completion on a voluntary basis. During the after-class interviews, students were asked brief questions about their learning experience related to the session and the peer assessment if applicable. In the end-of-course interviews, students were invited to give their opinions on the course during a more in-depth one-to-one interview session. The questions selected for the interview were closely related to the questionnaire because it was intended that the interview data would assist in giving a better understanding of the questionnaire results. Audio-recordings of the interviews were taken with the consent of the students. Each interview usually took 30 to 40 minutes to complete.

In both kinds of interview, the students were assured beforehand that the recordings or notes would only be used by the researcher for the purposes of the present research. The interviews were kept informal and casual in order to encourage the students to express their feelings.

3.6.5 Video Recording

Video recordings recorded the teachers' assessment interactions with students in the classroom. It was expected that the video recordings would provide the chance for the author to review interactional aspects of the teaching and learning as well as preserving valuable information on the design of the training. In addition, they also allowed details of the messages, intonation, body movement, and facial expression of participants during the interaction to be accurately documented, in order to analyse how the effectiveness of peer assessment could be achieved through a rather complex interaction process between the teacher and student

Due to practical limitations, recordings were only regularly made with class A's sessions. The camcorder was placed at the front of the classroom facing the students, so that many of the teacher's non-verbal expressions were not captured. Students were aware of the filming but reassured about its purpose. The camcorder was set up in such

a way as to minimize any distress that might be caused to the students by asking the students themselves to decide the exact position of the camera (within an acceptable range). It is believed that the associated anxiety would be gradually reduced when the recording is undertaken as a routine practice which the student will get used to (Allwright, 1991; Swann, 1994). Over time, all participants acted naturally as in a normal classroom.

By the end of the training, 15 hours of recording, including 7 hours of writing instruction and 8 hours of peer assessment training, had been collected. The 8 hours of peer assessment training was considered to represent a reasonable volume of data of interaction, comparing favourably with the 5-10 hours recommended by Seedhouse (1995).

3.7 Data Analysis

This section describes the tools utilised for analysing the data (see Table 3-4).

Table 3-4 Process of data analysis and interpretation

Research Sub-questions	Data Type	Data Analysis
What is the nature of peer feedback provided in each group?	Students' feedback/writing; questionnaire, interview.	Textual analysis; statistical analysis; theme analysis
How does the teacher model and negotiate to shape students' peer feedback?	Video recording	Discourse analysis

Details of the analysis methods used are given in the following sections.

3.7.1 Analysis of Students' Written Feedback

Two types of textual analysis were used with students' written feedback, the first of which is related to peer feedback. With reference to Min's (2005) method, the quality of reviewers' feedback was measured in terms of the number of 'steps' produced in each comment. Ideally, a comment would be considered most relevant and specific if it involved clarifying, identifying and expanding on a single issue with suggestions for improvement. For this reason, the author coded the single comments in brackets according to the prescribed criteria: identify a problem (identify step); explain the problem (explain step); clarifying writer's intentions (clarify step) and identifying specific suggestions (suggest step). These are shown in the examples below.

Example 1:

What kind of product do you have? And the target audience? Target region? The tone you used here is rather strong and too direct (**identify step**). Since you are writing as an employee to the upper level of the company, you should pay attention to the register you used in this mail (**explain step**). For example, a good company should have its own brand and product which it can target at all the customer groups so there is best commercial value for the company (**suggest step**).

Example 2:

‘If we don’t have it, we cannot map out a new plan with Mesa Food, and it including the work distribution and who has to be responsible for the deficit or the loss.’ I don’t understand what you try to say here. (**identify step**). Do you mean that Mesa Food should provide related information regarding the mentioned issue or do you mean that the concerns mentioned above are for Omni Incl? If they want to have a plan on it (**clarify step**)? I think you have a point on the argument but it will be more convincing if the relation of its cause-effect can be clearly stated (**explain step**). For example, if there is no information about the future plan of Mesa Food, it may lead to disagreement between both sides on the business principles and the lost benefits for buying the company (**suggest step**).

Min’s (2005) method of textual analysis helped in establishing the baseline for the present research. However, during the analysis it was found that this textual analysis method may be deficient. It was found that the suggestions made in a comment concerning an identified problem could be irrelevant. Hence, the textual analysis must consider the quality of the comment as well as its relevance to the problem identified.

Based on the above, a second, more holistic analysis was proposed to interpret the effects of the training and its overall quality, which is here defined as the specificity of the suggestion to the identified problem. A marking scheme was developed for this analysis (see Table 3-5). Two independent assessors analysed the comments and evaluated them using the agreed marking scheme. Examples of the grading can be found in Appendix F.

Table 3-5 Marking schemes of feedback and writing work quality

Mark	Relevance and specificity of formative peer feedback	Students’ writing work
5	Problems identified correctly completed with reason, completed valid suggestion	All reasons are valid complete with detailed explanation
4	Problems identified complete with reason, complete weak improvement	Most reasons are valid complete with detail explanation
3	Problems identified completed with reasons	At least 1 reason is valid

	complete with invalid suggestion	complete with detailed explanation
2	Problems identified without valid reason	At least 1 reason without detailed explanation
1	Irrelevant problems identified (subject to students writing work review)	No reasons are valid

The marking results were analysed using statistical analysis methods using the software suite SPSS 2.1. The following statistical methods were applied. The independent samples *t*-test is a test of the null hypothesis, and in this research it was used to compare the means of different groups in relation to the variation in the data (expressed as the standard deviation of the difference between the means). It was used to measure the differences in step usage in peer feedback as well as feedback quality between the two groups of students after each task. If the training was effective, the score for student's step usage and the quality of peer feedback in group A was expected to be significantly higher. The repeated measure of ANOVA was used to examine the effect of the second task of students' step use on the quality of formative peer feedback. This statistic allows researchers to monitor how participants change over the passage of time in a 2-by-2 research design (Field, 2000). In this research, in order to test the effects of the second task on peer reviewing ability, the two groups of students' usage of steps and the quality of peer feedback were tested.

3.7.2 Analysis of Students' Argumentative Writing Assignments

As introduced in section 3.6.2, in order to determine whether the peer assessment practice influenced students' writing quality, comparisons were made before and after the intervention. A score was assigned to the written work of each student using the criteria set out by the course requirements. These were independent of the peer feedback criteria, but due to the nature of argumentative writing, the criteria were very similar to the quality criteria detailed in Table 3-5. Hence, the author also adopted the same criteria for the writing quality evaluation. Examples of gradings are presented in Appendix G.

The result for the students' argumentative writing were analysed using the statistical program SPSS. An analysis of covariance (ANCOVA) and the paired sample *t*-test were employed to examine the effects of training on the quality of students' writing.

ANCOVA allows a comparison of one variable in 2 or more groups taking into account (or to correct for) other variables, called covariates. In this research, the quality of post-

training writing by students from the innovation and control group was compared (taking into account the relevant covariates) to their writing quality prior to the training. A paired sample *t*-test is a test of the null hypothesis used to compare the means of the same group in relation to variations in the data (expressed as the standard deviation of the difference between the means). In this research, the writing quality of students before and after peer assessment was measured. If the peer assessment was effective, the scores for students' writing were expected to be significantly higher after the peer assessment.

3.7.3 Analysis of Questionnaire Results

Students' perceptions and attitudes were investigated using two types of questionnaires: one before and one after the peer assessment practice, which are denoted as the pre-Q and post-Q respectively. Both types of questionnaires included 13 identical questions plus 3 additional questions in the post-questionnaire. Six aspects of students' perceptions and attitudes were considered in the questionnaires: attitude towards and perception of teacher/peer assessment; ability to produce peer feedback; perception of aspects of writing; peer collaboration; the use of peer feedback activity; and for additional questions in the post-questionnaire, the experience of peer assessment. The results from the pre and post questionnaires were analysed using the statistical program SPSS. Independent samples and paired samples *t*-tests were used. If the training in group A had an impact on students' perceptions and attitudes, the scores for the questions in group A were expected to be significantly different from those for the same questions in group B. In addition, the paired samples *t*-test was used to compare the groups before and after the training. If the training had an impact on students' attitudes and perceptions, then the scores generated in the post-questionnaire were expected to be higher. The results from the questionnaires were then compared with the quality of students' peer feedback and arguments using the ANCOVA. This aimed to identify the contribution of such perceptions towards the final outcomes of the experiment and also to identify any differences in behavioural patterns between the groups along with the different perceptions found in the questionnaire responses.

3.7.4 Recordings of Interviews

The interview data from sound recordings in Chinese was transcribed to support the results of the discourse analysis and the questionnaire results. The transcripts were then translated into English, and the translations were carefully checked by an expert

bilingual in Chinese and English. Four main themes were included to match the categories of the formative assessment (Pryor and Crossouard, 2010) and would referred to both teachers' and students' feedback. Information from the interviews is mainly verbal comments concerning the peer assessment training. The comments are referred to later in the thesis to support the quantitative results by providing detailed feedbacks from the students.

3.7.5 Analysis of Video Recordings

To conduct an analysis of observed interaction, it has been argued that it is crucial to provide a detailed transcription to allow for more accurate interpretation (Walsh, 2006). Therefore, the video recordings were transcribed following the methods used by Atkinson and Heritage (1984) to include details of word uttered, intonation, stresses, pauses, the overlapping of the participants' conversations and their behaviour (see Appendix H for a list of transcription conventions). Therefore, the transcription is a moment-to-moment record of the interactions. Both modelling and negotiation discourses were examined using Bernstein's (1996) concept of classification and framing which follows the suggestions of Morais (2002). An example is presented in Table 3-6.

Table 3-6 Description of classification

Case of classification	Description of the boundary between concepts/knowledge
C+	Knowledge of English argumentative writing is imparted through a strong argumentative concept and writing expression. The peer feedback steps are delivered in specific order and language.
C-	Knowledge of English argumentative writing is imparted through argumentative concept along with writing expression. The peer feedback steps are delivered in preferred order and language with problematic argumentative writing examples.

Upon examining the interaction transcription, four cases of regulative (RF) and instructional framing (IF) were found, based on which the relevant sentences were classified as shown in Table 3-7 (Morais, 2002):

Table 3-7 Description of regulative framing and instructional framing

Case of framing	Description of interactive relationships
------------------------	---

RF--	Students' participation and intervention is not controlled by the teacher. The emphasis is placed on the highest possible degree of students' participation in the interaction process. In this case, the students are asked to solve open-type problems or to work independently. This requires cognitive or social abilities that presuppose a high degree of autonomy on the part of the student.
RF-	Students' participation and intervention is not controlled entirely by the teacher. The emphasis is placed on the teacher's administrative role and a certain degree of student participation in the interaction process.
RF+	Students' participation and intervention is required in the act of teaching and/or the emphasis is placed on the teacher's administrative role.
RF++	The statements emphasis the teacher's guiding role. This illustrates a passive stance on the part of the students in the interactive process.
IF--	The teacher accepts students' answers with neutral attitude.
IF-	The teacher or students point out what is incorrect but do not correct the error.
IF+	The teacher or students point out what is correct / incorrect in a general way.
IF++	The teacher or students point out what is correct / incorrect in a clear and detailed way.

3.8 Reliability and Validity

Bryman (2007:28) specified four types of measurement validity: internal validity, external validity, and ecological validity.

In relation to measurement validity, the analytic tools used to examine students' work for step usage, quality of peer feedback and quality of argumentative writing, were and adapted from previous research (Min, 2005). The analytic tool used to examine classroom discourse was also adopted from a similar study (Morais, 2002). In both these examples, measurement validity was confirmed. In addition, the interview questions were piloted and checked to avoid leading questions. The interview data was analysed with the help of a type of thematic analysis used commonly in qualitative research. The validity of the questionnaire data could be tested statistically by examining variation in the results for each individual item in the questionnaire.

To ensure the internal validity of the observations, mixed methods and analytical approaches were used to collect and analyse different types of data to identify the major factors. The study had a single design but one control group was included. All of these factors enhanced the internal validity.

The external validity of this study is open to question. This is due to limitations relating to a small sample size and a sample of college students in their fourth year English writing class with lower levels of English proficiency. This study aimed to describe step

training practice, as the sample was expected to have included low, intermediate and high proficiency English learners in Taiwan.

The methodologies used in this research were specialized and generalization could only be achieved with careful modification for each specific class setting within the similar subjects.

3.9 Ethical Integrity

Ethics in social science is a difficult area. The focus in the literature has been on qualitative methods and on questions of confidentiality, informed consent and the minimization of harm (Berg, 2007). To inform the school authorities, the researcher provided a written handbook and gave a brief oral report about the study to the Director of the English programme and the class instructors. They all gave a warm welcome to the intervention. The participating students were informed of the study at the beginning of the course with a presentation and a letter of consent about: a) the purpose of the study; b) the issue of beneficence; c) the right to participate or withdraw from video shooting, interviews and questionnaire; d) a guarantee of no harm to them or disruption of their schedule; e) a guarantee of confidentiality, anonymity and non-traceability in the research; and f) a guarantee that data would be destroyed within three years after the PhD thesis was completed (see Appendix C). Although video recording was used, which strongly invades personal privacy, the study had no intention to look into personal affairs. Besides this, the participants were assured that the data would be kept and viewed only by the researcher, the co-rater in the process of analysis, and the research supervisors, and it would be used for academic purposes only. Therefore, no permanent harm could possibly be done to the participants. All interviewees were volunteers, and no complaints were made about any perceived unfairness or preferential treatment that these students would receive by agreeing to participate in the study.

Chapter 4 Results

4.1 Introduction

In Chapter 3, it was argued that the methodologies used in previous research into peer assessment were deficient, largely focusing on quantitative results concerning the effect of peer assessment training such as the quality of student's written work and the use of steps in student peer feedback. This study also investigates the effects of aspects of formative peer assessment training by gathering quantitative data on the use of steps and examining the quality of the feedback provided by the students. Those findings will help to answer the first research question 'What is the nature of peer feedback provided in both groups?' In previous research, qualitative results reflecting actual interaction during training has rarely been investigated in detail. In this study, therefore, the researcher sought to improve on previous studies by focusing on the details of interaction during the peer assessment training, such as by conducting discourse analysis. A quantitative analysis of details of the classroom interaction was used to investigate the alternation of framing within different discourses, leading to different reactions from the students. These results are then interpreted using the socio-cultural theory of formative peer assessment, which will help to answer the research question 'How does the teacher model and negotiate to shape students' peer feedback?' Moreover, questionnaires and interviews were used to investigate students' perceptions of and attitudes towards the peer assessment training, whereas the correlation of these factors with the effects of peer assessment training have not previously been investigated in detail. The results of these analyses were then used to gain a more comprehensive understanding of peer feedback training in an L2 writing classroom.

In this chapter, the results of the main teaching experiments are presented in two parts. The results presented in sections 4.2, 4.3 and 4.4 are mainly quantitative and include the analysis of background of the students, the effect of the formative assessment on the 'use of steps', a quantitative measure of their quality of the feedback and the quality of their written work. Interview and questionnaire results are also included to demonstrate the effects of the training on students' perceptions and attitudes. In line with previous research projects, these results provide a general evaluation of performance of the designed training activity. The results in section 4.5 are qualitative, and focus on the details of teacher/student interaction. The findings will help in understanding the processes underlying the effectiveness of the teaching and learning.

4.2 Background Analysis

The background analysis compared the level of student achievement in terms of the results of tests they had completed prior to the commencement of the experiment. The results confirmed that there are no significant differences between the two classes involved in the experiment.

A baseline was established from the past test results of the students in English proficiency test results from the past 4 years. In Table 4-1 it can be seen that the mean scores for the English language proficiency tests for the two classes were very similar during the four year period. Statistical analysis showed that the students in the two groups were at the same level of English language proficiency, with the independent *t*-test showing no significant differences between the test results for these 2 classes ($p > 0.05$).

Table 4- 1 Mean scores and independent *t*-test results for English language proficiency

	Group A		Group B		Independent <i>t</i> -test				
	Mean	Std. Deviation	Mean	Std. Deviation	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
Initial	140.8276	51.52759	128.3898	42.77286	1.044	.309	1.422	115	.158
Test1	137.7037	43.02243	129.1852	43.43899	.038	.846	1.024	106	.308
Test2	139.9615	62.84901	138.3889	56.23818	.554	.458	.136	104	.892
Test3	159.5417	43.01754	165.3462	49.55879	.979	.325	-.623	98	.535
Test4	161.9057	52.56507	166.1228	55.72672	.002	.968	-.408	108	.684

These results provide supporting evidence for the claim (see section 3.5) that the 2 classes were at a similar level of English proficiency. As introduced in section 3.5, the backgrounds of the two classes are similar as well, with both classes of final year undergraduate students majoring in Japanese.

4.3 Step Usage and the Quality of Students' Work

In this section, the quantitative measures of 'step usage' and the quality of students' written feedback during the experiment are reported. The experiment involved two tasks for peer feedback training: introducing criteria and negotiating criteria (see section 3.4.4). The step usage and the quality of the work produced by the two groups were measured with the methods shown in sections 3.6.1 and 3.6.2.

4.3.1 Statistical Methods

The independent pair *t*-test was used to examine the step usage and quality of peer feedback of the 2 groups. The independent pair *t*-test was used to compare the means of different groups in relation to the variation in the data (expressed as the standard deviation of the difference between the means). In this study, this statistic was used to investigate the difference in step usage in peer feedback as well as quality, which was quantified according to the marking scheme so as to be able to analyze it using statistical tests between the two groups of students after each task. If the training was effective, the score for students' step usage and quality of peer feedback in group A (which followed Min's step training method) was expected to be significantly higher.

In addition, one way analysis of covariance (ANCOVA) and the paired sample *t*-test were also employed to examine the effects of training on the quality of students' writing. ANCOVA allows the comparison between two or more groups for one variable taking into account (or to correct for) are variability of other variables, called covariates. In this study, the quality of the writing produced by students in the experimental and control groups after training were compared, taking into account the students' pretest results.

4.3.2 The 'Use of Steps' in Students' Formative Peer Feedback

One of the primary measures of the effectiveness of peer feedback training was the use of 'steps' in the students' comments. These 'steps' were introduced by Min (2005:303) to achieve effective peer feedback on the basis that *"reviewers' feedback was much more relevant and specific because it revolved around clarifying, identifying and expounding a single issue and suggesting a way to improve it."* Despite a few drawbacks, the 'steps' approach does have a number of advantages, and a detailed review of this method can be found in section 3.7.1. The author first followed Min's approach to measure the effectiveness of the steps in the students' peer feedbacks; the methodology of the experiment is detailed in the previous chapter (see section 3.4). Unlike in Min's (2005) work, students in this study produced only one comment per peer assessment, from which the different steps were identified and calculated as a score under each step category (see section 3.7.1). As defined by Min, there are four steps which can be identified within student's peer feedback. The quality of the peer feedback is considered to be proportional to the number of steps contained in the feedback. More

step usage represents higher feedback quality. The average number of steps used by the students in a class is then calculated.

The results shown in Table 4-2 show the use of each step in both groups and both tasks, and the independent *t*-test outcomes demonstrate the differences between the groups. Group A, the experimental group, made more use of almost all four steps in feedback than group B did in the first task. Aside from this divergence, both groups produced similar results.

Of the four steps, the usage of ‘clarify’ step was much lower than others, indicating that the training sessions in both groups had provided less guidance to students on how to use this step compared to other steps. The author will discuss the implications of this later (see section 5.1.2.2). There were significant differences between the two groups in their average usage of the ‘clarify’ step after both tasks 1 and 2 ($p < 0.05$). This indicates that the difference between the training in group A and group B had influenced the usage of this step.

‘Identify’ was the step most used by students in both groups. However, there is no significant difference between groups A and B on using this step after task 1, suggesting that the training sessions in both groups may have provided effective help to students in using this step. Conversely, after task 2, group A’s level of the usage of this step diverged from that of group B, indicating variations in the impact of the training during task 2 which results in group A subsequently outperforming group B in terms of using this step. Considering the definition of this step, one may deduce that the step training for group A was more effective in terms of helping students to identify problems during peer assessment.

There were significant differences between the two groups in the use of the ‘explain’ step and ‘suggest’ steps during both tasks, again suggesting that the differences in training may have significantly impacted on the two groups’ usage of these steps. For both tasks, group A outperformed group B in terms of using these two steps. Based on the definition of these two steps, the results may lead to the conclusion that the step training for group A helped the students to explain the problem identified in the peer assessment and to suggest corrective action in their peer assessment feedback.

In Table 4-2 the mean scores of step usage represents the average values of each step used by the group of students. The maximum value for the mean score is 1, which

represents every student in the group using the step once. A higher mean value of the step represents more usage and *vice versa*. For example, during task 1, group A students used the clarify step 0.1111 times on average. This represents only a small proportion of students in group A using this step. During task 1, group B students used the step ‘identify’ 0.9245 times on average. This represents this step being used by most of the students.

Table 4-2 Mean scores and independent *t*-test result for step usage

Steps	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	t	df	Sig. (2-tailed)
Clarify_task 1	.1111	.31722	.0000	.00000	33.965	.000	2.550	105	.012
Clarify_task 2	.1132	.31988	.0185	.13608	18.547	.000	1.999	105	.048
Identify_task 1	.9259	.26435	.9245	.26668	.003	.957	.027	104.92	.978
Identify_task 2	.9434	.23330	.6481	.48203	103.628	.000	4.021	105	.000
Explain_task 1	.6296	.48744	.1509	.36142	29.817	.000	5.762	105	.000
Explain_task 2	.7358	.44510	.0370	.19063	71.760	.000	10.591	105	.000
Suggest_task 1	.6111	.49208	.1887	.39500	21.263	.000	4.891	105	.000
Suggest_task 2	.7547	.43437	.1852	.39210	2.278	.134	7.115	103.49	.000

Table 4-3 ANCOVA results comparing the use of steps by the two groups

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.
Task * Group	Clarify	.005	1	.005	.073	.788
	Identify	1.046	1	1.046	10.030	.002
	Explain	.580	1	.580	4.251	.042
	Suggest	.245	1	.245	1.467	.229

In Table 4-3, the ANCOVA results show whether or not step usage is influenced by the difference between peer assessment tasks. Significant differences were shown in the usage of the ‘identify’ ($F(1,105)= 10.030, p<0.05$) and ‘explain’ steps ($F(1,105)= 4.251, p<0.05$) between tasks. The variations were affected by the decrease in the use of these two steps in group B during task 2 (Table 4-2), highlighting the impact of the difference between tasks on students’ usage of these steps. Group B students used ‘Identify’ and ‘Explain’ significantly less in Task 2 than they did in Task 1, and also much less than Group A in Task 2. Group B students demonstrated similar levels of step usage in relation to the ‘clarify’ steps and ‘suggest’ steps, indicating that their capacity and willingness to annotate these steps was not affected by the change of task. In the meantime, Group A students’ usage of all four steps increased from Task 1 to Task 2.

This suggests that the step training had a positive effect on group A's use of these steps in the second task, although their level of use was already high in the first task. The drop of step usage in group B demonstrated that, without the step training, the students in group B could not produce certain steps as well as they did in Task 1 during which they had been given clear instructions on how to make peer assessment feedback. This again demonstrated the positive effects of the step training, which may lead to the conclusion that the step training can help the students to form a solid memory about step usage during peer feedback, which may then improve the quality of their feedback.

4.3.3 The Quality of Students' Formative Peer Feedback

The quality of peer feedback is evaluated using the coding system shown in section 3.7.1, and is determined by its relevance to and how specific the suggestion made was concerning the problems identified. Min's (2005) coding system was used to quantify the quality of students' work.

In Table 4-4, the mean score of the quality of peer feedback is the average mark of the groups using the criteria and coding system defined in 3.7.1. The highest mark for feedback quality is 5. A higher mark represents a higher quality of the feedback.

Table 4-4 Mean scores and independent *t*-test results for the quality of peer feedback

Task	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Deviation	Mean	Std. Deviation	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
1	3.29	1.261	.98	.537	46.618	.000	12.292	107	.000
2	3.0	1.202	1.13	.621	13.255	.000	10.470	111	.000

Table 4-5 ANOVA results for the quality of peer feedback between tasks

Effect		Value	F	Hypothesis df	Error df	Sig.
Task* Group	Pillai's Trace	.025	2.796a	1.000	107.000	.097

The effects of training on the usage of individual steps have been summarized in the previous section. In this section, data on the quality of student feedback is provided. The figures in Table 4-4 compare the quality scores between groups A and B. In order to clarify the definition of the marking quality, it is useful to revisit the explanation in section 3.7.1. Reference was made to Min's (2005) method, where the quality of reviewers' feedback was measured by the number of 'steps' produced in each comment. Ideally, a comment would be considered most relevant and specific if it involved clarifying, identifying and expanding on a single issue with suggestions for improvement. For this reason, the single comments were coded in brackets according

to the prescribed criteria: identifying a problem (identify step); explaining the problem (explain step); clarifying the writers' intentions (clarify step) and identifying specific suggestions (suggest step). Min's (2005) textual analysis method helped the author to establish the basis for this research. However, during the analysis it was found that the text analysis method may be deficient, because the suggestions made in a comment for an identified problem could be irrelevant in a comment even if all of the steps were all present. For example, see the following feedback:

I think the arrange is too small that's can't get more profit for us, because 14-25 year old is younger we should to open more customer arrange Why do you think the range is too small. It is good to have a range and target on this group of people with special designed products. If you think there is a need to increase the age range, can you suggest how? There is a need to understand the potential of the products for different age group such as baby food or food for elder adults. Also, is there facility for the study and invention.

In the above feedback, despite the deficiencies in language (the student also orally explained the feedback in Chinese, and hence the comments can be considered to have been completed to an acceptable standard during the session), the student seems to complete all four steps. However, if we look closely at his expression, it is clear that his suggestion in the later part of the feedback is not relevant to his explanation in the earlier part of the feedback. Hence, his suggestion is not based on his reasoning; instead, it was rather random. This kind of feedback should not be considered as effective in terms of a valid suggestion.

Given the above observation, it was decided that the textual analysis must consider not only the step usage but also the relevance to the problem identified. Based on this, an improved, more thorough analysis was used. The marking scheme based on the improved criteria is as shown in Table 3-5 which is reproduced below:

Table 3-8 Marking schemes of feedback and writing work quality

Mark	Relevance and specificity of formative peer feedback	Student writing work
5	Problems identified correctly complete with reason, complete with valid suggestion	All reasons are valid complete with detailed explanation
4	Problems identified complete with reason, complete with weak improvement	Most reasons are valid complete with detailed explanation
3	Problems identified complete with reasons,	At least 1 reason is valid

	complete with invalid suggestion	complete with detailed explanation
2	Problems identified without valid reason	At least 1 reason without detailed explanation
1	Irrelevant problems identified (subject to students writing work review)	No reasons are valid

The student feedbacks was then marked based on the above criteria,. Examples of marking are shown in the following table (the table is not numbered as is identical in Appendix F, shown here for clarification purpose only):

Mark	Relevance and Specificity of formative peer feedback	Students' examples
5	Problems identified correctly complete with reason, complete with valid suggestion	You should point out which company we should know more about and which company will benefit from the investment. If you can point that out clearly in the writing, it will be easier for readers. For example, we need to understand more about Mesa Foods such as their products, the strength and weakness, so it will be better to our company, Omni Inc.
4	Problems identified complete with reason, complete with weak improvement	I think there is no relations between the age of target group and the decision of buying Mesa Food since 14-25 is the target customer of Omi Incl. You can say instead that snck food are popular among children or young group which is the target group of Omni Incl, but we don't understand their way of spending money neither the allowance for the cost. In other words, you should explain first what is the target customer of Mesa Food and their ability to spend and consume on such product.
3	Problems identified complete with reasons complete with invalid suggestion	I think the arrange is too small that's can't get more profit for us, because 14-25 year old is younger we should to open more customer arrange Why do you think the range is too small. It is good to have a range and target on this group of people with special designed products. If you think there is a need to increase the age range, can you suggest how? There is a need to understand the potential of the products for different age group such as baby food or food for elder adults. Also, is there facility for the study and invention.
2	Problems identified	Market analysis, you can give more examples on

	without valid reason	that. Revise on reason 3, you can argue that the target customer is different in two companies. You can try to write in this way: Since the target group is different, there is a need to understand the needs of the target group before any decision is made.
1	Irrelevant problems identified (subject to students writing work review)	‘The snack more delicious and more attractive,’ how can you make the snack more delicious and more attractive?

It is shown that group A produced better quality peer feedback writing than group B in both tasks. The mean scores of group A are greater than those of group B in both tasks and the t-tests confirm that the variation between the two groups is significant ($p < 0.05$). In Table 4-5, the ANOVA results suggest that there was no significant effect of the task change on group A and B's quality of peer feedback ($F(1,107)=2.796$, $p > 0.05$). The results above indicate a clear advantage was gained by group A over group B by their use of the 'steps' training. The next question is whether or not the improved feedback quality had an impact on students' written work.

4.3.4 The Quality of Students' Written Assignments

In the previous section, the examination of both step usage and the quality of peer feedback has suggested the positive effects of the steps training. In this section the impact of the improved quality of feedback on student's written work is presented.

As introduced in 3.6.2, in order to determine whether the peer assessment practice influenced students' writing quality, before and after comparisons were made. The pre-test results represent the quality of the students' writing assignments quality before the peer assessment, and the post-test results represent their writing quality after the peer assessment. A score was assigned to the written work of each student as per the criteria set out in the course requirements. These are independent of the peer feedback criteria, but due to the nature of argumentative writing, the criteria were very similar to the quality criteria detailed in Table 3-5. Hence, the same criteria were also adopted for the writing quality evaluation. Examples of grading are presented in Appendix G.

In Table 4-6, the 'pre-test' and 'post-test' results of each group are compared. The improvement in pre- and post-test marks for group A is significant ($t=-4.138$, $p < 0.05$) while group B shows no significant improvement ($t=1.786$, $p > 0.05$). This suggests that the peer assessment had different impacts on the writing of the two groups.

Table 4-6 Paired *t*-test results for the quality of students' writing before and after training

Group		Paired Differences					<i>t</i>	df	Sig. (2-tailed)
		Mean	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
A	Pre-test – post-test	-.728	1.303	.173	-1.074	-.382	-4.220	56	.000
B	Pre-test – post-test	-.396	1.714	.235	-.869	.076	-1.683	52	.098

Table 4-7 Mean scores and independent *t*-test result for the quality of students' writing

Test	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
Pre-test	2.64	1.063	2.06	1.106	.053	.818	2.878	109.182	.005
Post-test	3.38	1.060	2.43	1.248	2.972	.088	4.342	110	.000

Table 4-7 also shows that students in group A showed significantly better writing quality than group B both before and after the peer assessment. It was therefore necessary to make sure that the difference between the two groups in the pre-test results did not have a significant impact on the post-test results. To clarify this, an ANCOVA analysis was performed. This assumed that the pre-test results represent the disparity between the two groups and that the writing quality score is a variable dependent on two covariates: the pre-test results and the feedback quality.

Table 4-8 ANCOVA results for the quality of students' writing

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Pre-Test	1.459	1	1.459	1.137	.289	.010
Feedback Quality	6.142	1	6.142	4.786	.031	.042

To interpret the ANCOVA results, the following values should be focused on. 1. The *F* values represent the ratio of variance between groups over the variance within a group. The bigger the *F* value the more difference the covariate will make. 2. The significance value (or *p* value) is a measure of the significance of the influence on the results. In this study, $p < 0.05$ is considered to be statistically significant. 3. The partial eta squared value, which represents the relative importance of one covariate compared to another one. The bigger the partial eta squared value the more contribution this covariate will make towards the result.

Table 4-8 shows the ANCOVA analysis results showing the influence on the post-test writing quality of the two covariates, the pre-test result and the feedback quality. The significance figure (Sig.) allows us to decide if the covariate has a significant contribution to the post-test results. The table shows that the pre-test results did not have

a significant influence on the post-test results. The contribution of this variable is much lower than that of the feedback quality (pre-test, $F=1.137$, $p>0.05$). On the other hand, the influence of feedback quality is significant ($F=4.786$, $p<0.05$) and the impact is much greater (the value of partial eta squared for feedback quality is 0.042, which is more than four times the value for the pre-test figure (0.010)). The ANCOVA results prove that feedback quality had a significant impact on the quality of students' writing.

4.3.5 Findings

In this section, the results of quantitative analysis based on the use of steps, the quality of student's peer feedback and writing work are presented.

(1) Step usage in peer feedback

The experimental group utilized more steps than the control group, which made similar use of the step 'identify' but significantly less use of the other steps. This is comparable to Min's (2005) results and confirms that training which focuses on step usage significantly improves the quality of the students' work.

(2) The quality of peer feedback

The experimental group's quality of peer feedback was better than that of the control group. Marked by the criteria set out in section 3.7.1, the peer feedback quality was evaluated. The results further demonstrated that the quality of the peer feedback was improved as a result of the 'steps' training. The evidence lends support to Min's claim that step training can improve the quality of peer feedback.

(3) The quality of students' writing assignments

The experimental group's writing quality was also better than the control group. This result may lead to the conclusion that the final product of peer assessment in this study – the writing quality – was improved as a result of the improved quality of peer feedback. The group with the better peer feedback quality shown a significant advantage when it came to making improvements and revisions to their written works. However, the quantitative results cannot fully explain the reasons for the positive outcomes and are not fully convincing due to many practical reasons, including the lack of strict control over the quasi-experimental environment, the lack of precise quality measurements and the absence of repeatable measures.

4.4 Students' Opinions, Attitudes and Perceptions

Apart from the writing and feedback results, the author also deployed questionnaires to elicit students' opinions of and attitudes and perceptions towards the experiment. These measures are important indicators of the effects of peer assessment training on students. Moreover, the author also conducted a series of interviews with students to investigate further the students' opinions of, and attitudes and perceptions towards, the experiment.

As explained in previous chapters, the students' perceptions, attitudes and opinions could have had an influence on the final outcome of the experiment. To examine these potential influences, measurements of these psychological aspects are analyzed alongside the quantitative results from the previous section. Questionnaires (see Appendix E) were issued to both groups before and after the experiment containing two sets of questions. The responses were graded from 1 to 5, where 1 indicates the least agreement with the given statement and 5 the highest agreement with the statement.

For convenience, the following abbreviation are used:

Questionnaires issued before the experiment are termed: pre-Q.

Questionnaires issued after the experiment are termed: post-Q.

The questions are referred to as follows:

Question 1 in the pre-Q is termed: preQ1.

Question 1 in the post-Q is termed: postQ1.

Note also that questions 14-16 only appeared in the post-peer assessment questionnaire.

The questionnaire data were analyzed by comparing the mean response values for each question between groups and tests. The comparison is made to find the general trends and any obvious differences in psychological factors between groups or at different times. It is obvious that the scales used in the questionnaire measurement at neither uniform nor precise. Hence, the results for different psychological factors will be compared with each other. Independent and paired *t*-tests were used to determine the significance of variations between results. ANCOVA is then used to examine the relationship between the measurements and the experimental outcomes. Interview data was gathered from students randomly selected from both groups, and two types of

interview were conducted. The first interviews were carried out with 4 students (who were available after being randomly selected) before and during the experiment, and the second set were conducted with a further 2 students two weeks after the completion of the experiment. The interview questions were generally adapted from the questionnaire items, and details can be found in Appendix D. A thematic analysis of the data was carried out, using themes selected from the following categories which are of particular interest in this study:

- Teacher or peer assessment
- Ability to produce peer feedback
- Writing with peer assessment
- Experience of peer collaboration in peer feedback activity
- Experience of step training

The themes are then narrowed down to the actual issues expressed by the students during the interviews. The frequency of students raising each theme is then calculated to show its importance.

4.4.1 Students' Perception of their Ability to Produce Peer Feedback

As shown in Table 4-9, question Q7 asked the students if they thought they had the ability to peer review the work of others. Before the experiment, students from both groups did not consider themselves to have this ability (M= 2.6897 for group A and M=2.3208 for group B). After the experiment, both groups were asked the same question. The results shows very little improvement for either group (M= 2.7241 for group A and M=2.5094 for group B after experiment).

Table 4-9 Mean scores and independent *t*-test results for Q7

Questions	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
PreQ7	2.7000	.78762	2.3208	1.07011	5.309	.023	2.207	119	.029
PostQ7	2.7167	.66617	2.5094	.89061	6.588	.012	1.410	111	.161

As shown in Table 4-10, the paired *t*-test of pre- and post-test results also showed no significant variation in either group. Two themes were raised by the students during the

interviews. Most students from group A (6/6) did not consider themselves competent in peer reviewing the work of others. When asked for the reason for this, most responses (6/6) expressed the concern that the students were not confident about their English proficiency levels. Students from group B held similar views towards their peer assessment capabilities for similar reasons. One student (1/6) raised an issue regarding the difficulty of the peer assessment tasks and suggested that they should have started the peer assessment practice at the beginning of their English study with easier materials instead of working on more difficult tasks at this later stage. This feedback reflected the fact that some students believe that peer assessment is only achievable with more advanced levels of English proficiency. Apart from the concern about the language proficiency, some students pointed out other issues which led to their evaluation of their ability in peer assessment. One student said that: *'Due to the different thinking level between me and my peer, my comments are not necessarily understandable for my peer.'* The thinking level referred to by the student was explained as due to the fact that members of the class had different ability levels, and hence some of them may have had different understandings of the peer assessment training. In this case, the pair would not be able to work as peers since their understanding of the peer assessment differs.

Table 4-10 Mean variations and paired *t*-test results for Q7

Group	Mean Difference (Pre - Post)	Std. Dev.	Paired <i>t</i> -test result		
			<i>t</i>	df	Sig. (2-tailed)
A	-.01667	1.09686	-.118	59	.907
B	-.18868	1.27183	-1.080	52	.285

4.4.2 Students' Perception of Working on Peer Feedback

Questions 8, 9 and 10 asked about the students' perception of the focus of effort during peer assessment.

8. *My peer feedback focuses on grammar correction.*

9. *My peer feedback focuses on content.*

10. *My peer feedback focuses on organization.*

Table 4-11 shows that group A and B students had different perceptions ($p < 0.05$) of the focus on grammar issues in the peer feedback. Group A showed less agreement with statement 8, which means they might like to focus less on grammar correction than group B. There are no significant differences between the groups in attitudes towards the aspects mentioned in statements 9 and 10, which were content and structure.

Table 4-11 Mean scores and independent *t*-test results for Q8, 9 and 10

Questions	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
PreQ8	2.8333	.88618	3.4340	1.06535	3.043	.084	-2.929	113.722	.004
PostQ8	2.6833	.81286	3.1132	1.08593	5.255	.024	-2.399	111	.018
PreQ9	3.3667	.71228	3.3208	.95623	5.525	.020	.286	119	.775
PostQ9	3.4333	.76727	3.4906	.82328	.393	.532	-.381	106.927	.704
PreQ10	3.2167	.80447	3.1132	.93357	2.068	.153	.601	114.547	.549
PostQ10	3.2833	.78312	3.2075	.92733	.655	.420	.466	102.353	.642

Table 4-12 shows that although both groups of students had a tendency to reduce their focus on grammar, since the means for groups A and B were both lower after the peer assessment practice, these decreases were not statistically significant ($p > 0.05$). Group B had relatively (not absolutely since $p > 0.05$) more statistically significant ($p = 0.101$) changes than group A. On the other hand, both groups were of the perception that they should have paid more attention to content and structure. This was reflected in mean values shown in table 4-11. However, such changes were again not statistically significant ($p > 0.05$).

Table 4-12 Mean variations and paired *t*-test results for Q8,9 and 10

Question	Group	Mean Difference (Pre - Post)	Std. Dev.	Paired <i>t</i> -test result		
				<i>t</i>	df	Sig. (2-tailed)
8	A	.15000	.95358	1.218	59	.228
	B	.32075	1.39744	1.671	52	.101
9	A	-.06667	1.05552	-.489	59	.626
	B	-.16981	1.31172	-.942	52	.350
10	A	-.06667	1.11791	-.462	59	.646
	B	-.09434	1.18101	-.582	52	.563

The training in group A had strongly advised students to focus on the more general aspects of writing (content and structure) rather than grammar in their peer feedback. In the follow-up interviews, however, students reflected on the difficulties that this advice

had caused. During the interviews, comments like the following were raised: *‘At the first draft, I wrote without logic and sometimes off the topic. This training however, makes this class not an English class but a logic class. I want more language or grammar training’*; *‘my English is not good so I cannot give good suggestion’*. Class B students also reflected similar concerns, stressing that: *‘Due to the poor grammar and general skills of writing, the peer assessment cannot be effective’*. Similar concerns in the interviews are revealed in the following section. Students were not highly proficient with English writing and this shortage of skills surely hampered their efforts at providing peer feedback. Consequently, they generally lacked confidence in their written work and peer feedback. These reflections also showed that the students desired grammar training as part of their writing practice. During the interviews, among a number of comments about the training and the practice of peer assessment, many raised similar concerns. This reflects the fact that the students’ learning needs were not satisfied. The peer assessment training was not a practice they were familiar with and they felt let down by the lack of conventional advice. This raises the question about whether peer assessment training and practice with students of lower proficiency is worthwhile. Even with the step training, students still felt not fully prepared. What else do the teachers need to do to solve this problem?

4.4.3 Students’ Perceptions of Writing

In answer to question Q16 (which stated: I considered myself to be capable of producing English writing), students responded that they did not feel any more capable in English writing after the peer assessment ($M=2.18$ for group A, 2.298 for group B) as shown in Table 4-13. There was no significant difference between the two groups ($p>0.05$).

In a follow-up interview, 3 out of 6 participants in group A replied that grammar is still the most important aspect for them in good English writing. The students reflected that, *‘if you cannot write well in English with correct grammar, it is difficult to express yourself correctly, no matter how solid and strong the point was.’* Students in group B generally agreed with these comments during the interviews and members of both groups said they still lacked confidence in their English proficiency. This fundamental concern undermined any advanced activities in the classroom and also meant that students’ attitudes towards writing stayed within a narrow concept of peer assessment feedback.

Table 4-13 Mean scores and independent *t*-test results for Q16

Questions	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
PostQ16	2.1803	.64571	2.2982	.82299	7.048	.009	-.869	116	.387

4.4.4 Students' Attitudes towards Peer Assessment Training

Question Q6 asked the students if their training had helped them to think and to provide comments. As shown in Table 4-14, both groups agreed that their training had been of use (Mean > 3) before and after the experiment, although only group A received the 'steps' training. As shown in Table 4-15, the results also show increased mean values in both groups after the experiment: Group A (3.3167 to 3.45, $p > 0.05$), Group B (3.1509 to 3.5094, $p = 0.050$). The impact on their perceptions of the value of the training was significant. Group B students did not receive the 'steps' training but adopted a traditional approach as described in section 3.4.3 and 3.4.4.

Table 4-14 Mean scores and independent *t*-test results for Q6

Questions	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
PreQ6	3.3167	.65073	3.1509	1.02650	9.527	.003	.951	119	.343
PostQ6	3.4500	.87188	3.5094	.89061	.046	.830	-.358	108.670	.721

Table 4-15 Mean variations and paired *t*-test results for Q6

Group	Mean Difference (Pre - Post)	Std. Dev.	Paired <i>t</i> -test result		
			<i>t</i>	df	Sig. (2-tailed)
A	-.13333	1.15666	-.893	59	.376
B	-.35849	1.30228	-2.004	52	.050

After the 'steps' training, the students in group A provided specific comments. On examination, the comments in some cases were copied from the teacher's examples. When asked about this, 3 out of 6 students admitted that due to pressures of time they would copy the teacher's answers or produce something similar. This could explain some of the feedback described in section 4.3.3 which contained the steps but did not

contain valid suggestions. The training not only helped students produce systematic peer feedback structured in terms of the ‘the steps’, but 4 out of 6 interview participants commented that it also assisted them in thinking better about their own work. Students in group B also reflected on the positive effects of peer assessment practice and 2 out of 6 on how the peer assessment practice had helped them to think more critically. However, they also reflected that the practice may have focused too much on structural instead of grammar issues and because of this it may not directly lead to an improvement in their English proficiency. It is interesting that both groups agreed that the training in peer assessment had helped them with generally positive attitudes shown towards the trainings, despite the fact that the two classes had different approaches in term of trainings and that the quality of peer assessment was different.

4.4.5 Students’ Opinions of Time Allocated to Peer Assessment

Question 14 asked students about the time allocated to peer feedback activity during the class (statement: the time allocated for peer feedback activity is not enough). Group A agreed (Mean=3.2) that the time was insufficient, although group B disagreed (Mean=2.8). As shown in Table 4-16, this variation is significant ($p<0.05$) and reflects the fact that the students who participated in the ‘steps’ training would have preferred more time to be made available. However, the students in group B thought that the peer assessment activity had been allocated plenty of time.

Table 4-16 Mean scores and independent *t*-test results for Q14

Questions	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev	Mean	Std. Dev	F	Sig.	<i>t</i>	Df	Sig. (2-tailed)
PostQ14	3.2000	.81926	2.8302	.80230	.004	.949	2.421	109.804	.017

In follow-up interviews, 5 out of 6 participants from group A reflected on the ‘lack of time’ issue, and comments included that: ‘*Peer assessment should be given more time*’ and ‘*I do not have enough time to think before giving comments*’. These responses indicate that ‘steps’ usage was a time-consuming process and that the time remaining was insufficient, or that the students were feeling stressed and not given a chance to think properly. In group B, on the other hand, the students felt that the time allocated to the peer assessment without steps training was sufficient and even, in some cases, too much. One student commented that ‘*for the rest of the time in the class, I do not know*

what to do'. These results indicate that the arrangements for allocating time could be improved. However, the underlying issue of time allocation in group A is more fundamental. Group B's responses suggest another issue: that their teacher provided the students with sufficient time but not enough guidance. Detailed discussions of these issues are included in the following chapter.

4.4.6 Students' Opinions of Peer Collaboration in Peer Feedback Activity

Question 11 dealt with general opinions on peer collaboration (statement: I cooperate well with my peers in peer review activity), while question 12 considered the students' state of mind during the interaction (I am frustrated when taking part peer review), and question 13 sought the students' opinions on the level of conflict occurring during the interaction (statement: I feel offended when doing peer review). For question 11, both groups generally gave positive responses. As shown in Table 4-18, Group A students' scores for their level of collaboration were reduced after peer feedback practice. This change was significantly different from group B's evaluation ($p < 0.05$). Their level of collaboration increased slightly. For question 12, both groups gave negative feedback. As they both felt less frustrated. However, group A students were significantly less frustrated than those in group B ($p < 0.05$) both before and after the activity (see Table 4-17). Question 13 revealed different tendencies in the two groups. Table 4-18 shows that group A demonstrated increased levels of conflict, while in group B reduced levels. Question 13 answered before the peer assessment session also revealed that students in group A and B had different opinions before the peer feedback sessions took place. Group A students presumed there would be little conflict but in the end the conflict level increased. Group B students presumed a higher level of conflict but after the practice their expectation of conflict had reduced significantly ($p < 0.05$). This tendency is consistent with that shown in response to question 11.

Table 4-18 Means difference and paired t-test results for Q11, 12 and 13

Question	Group	Mean Difference (Pre - Post)	Std. Deviation	Paired <i>t</i> -test result		
				<i>t</i>	df	Sig. (2-tailed)
11	A	.20000	.89821	1.725	59	.090
	B	-.07547	1.26869	-.433	52	.667
12	A	.16949	1.17668	1.106	58	.273
	B	.30189	1.21821	1.804	52	.077
13	A	-.06780	.99766	-.522	58	.604
	B	.52830	1.29502	2.970	52	.004

As shown in Table 4-18, except for group B's responses to question 13, all the other answers failed to reveal significant differences between the pre and post-questionnaire data. During the interviews, students from group A reflected on conflict during peer reviews: *'sometimes, you don't agree to what your peers point out'; 'I understood what he said but his way of expression can be confusing and I found it hard to accept his point'; 'I would argue with his points because of the way he put it. When we have that, I would say no more because it would not work even if I tried to convince my partner'*. The latter student was obviously disappointed and had been discouraged by the conflict. Another student suggested: *'I think the pairs should have similar attitudes and the same level of criticality. For example, if a partner is over-critical; the pair would never work together to produce a workable peer feedback.'* The student here was concerned about the overly critical behaviour of his/her peer, and the reaction reflected a desire for a more 'workable' approach.

4.4.7 Students' Attitudes towards the Teacher's Role in Assessment

Questions 1 and 4 asked about students' attitudes towards the role of teachers in the peer assessment. Question 1 asked whether the teacher 'should be the reader and evaluator of my writing.' After the experiment, both groups agreed with this statement. Group A initially had a negative overall attitude towards this point before the experiment. But, as shown in Table 4-20, this changed significantly after the experiment (Mean=2.56 for preQ1, Mean=3.95 for postQ1). Group B initially (Mean=3.96 for preQ1, Mean=4.3 for postQ1) agreed with this question, and their level of agreement also increased at the end of the experiment. This clearly demonstrates that, at the end of the experiment, both groups were more in favour of the teacher being in charge of reading and evaluating their work. Question 4 asked students to consider if 'Teacher assessment can be replaced by peer feedback activity'. Students in both groups became more negative about this issue. In group A, the change was not significant, whereas in group B disagreement became more pronounced. These results indicate that, after the experiment, the students preferred more teacher assessment rather than peer assessment.

Table 4-19 Mean scores and independent *t*-test results for Q1 and 4

Questions	Group A	Group B	Independent <i>t</i> -test result
-----------	---------	---------	-----------------------------------

	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
PreQ1	2.5593	1.31662	3.9623	1.09126	5.401	.022	-6.522	119	.000
PostQ1	3.9492	.77512	4.3019	.63805	.142	.707	-2.639	109.199	.010
PreQ4	2.6167	1.00998	2.5094	1.06740	.264	.609	.643	117.673	.521
PostQ4	2.8833	.66617	2.5283	.82284	6.321	.013	2.533	111	.013

Table 4-20 Mean differences and paired *t*-test results for Q1 and 4

Question	Group	Mean Difference (Pre - Post)	Std. Deviation	paired <i>t</i> -test result		
				<i>t</i>	df	Sig. (2-tailed)
1	A	-1.38983	1.52017	-7.023	58	.000
	B	-.33962	1.27012	-1.947	52	.057
4	A	.11667	1.16578	.775	59	.441
	B	.30189	.86791	2.532	52	.014

4.4.8 Students' Evaluation of Motivation

Questions 2 and 3 asked about motivation in the class. Question 2 asked the students to evaluate the statement: 'My teacher is good at motivating'. Both groups agreed that the teacher had motivated them. In Table 4-21, Group B students (Mean=4.00 for preQ2, Mean=3.7 for postQ2) evaluated more highly than those in group A (Mean=3.6034 for preQ2, Mean=3.45 for postQ2), the variations are shown in the first and second rows of Table 4-21. These results show that group B students had higher perceptions of the motivational abilities of their teachers than did those in group A. Question 3 asked the students to evaluate the statement: 'My peer motivates me'. Both groups agreed that their peers had motivated them. In Table 4-21, Group B students (Mean=3.16 for preQ3, Mean=3.43 for postQ3) evaluated this more highly than those in group A (Mean=3.16 for preQ3, Mean=3.23 for postQ3), although the variations are not statistically significant.

Table 4-21 Mean scores and independent *t*-test results for Q2 and 3

Questions	Group A		Group B		Independent <i>t</i> -test result				
	Mean	Std. Dev.	Mean	Std. Dev.	F	Sig.	<i>t</i>	df	Sig. (2-tailed)
PreQ2	3.5500	.98161	4.0000	.70711	10.997	.001	-2.755	119	.007
PostQ2	3.4333	.78905	3.6981	.60717	4.203	.043	-1.979	111	.050
PreQ3	3.1667	.80605	3.1698	.82592	.085	.771	.204	118.220	.839
PostQ3	3.2333	.83090	3.4340	.57209	2.854	.094	-1.509	105.005	.134

Table 4-22 Mean differences and paired *t*-test results for Q2 and 3

Question	Group	Mean Difference	Std. Dev.	Paired <i>t</i> -test result
----------	-------	-----------------	-----------	------------------------------

		(Pre - Post)		<i>t</i>	df	Sig. (2-tailed)
2	A	.11667	1.16578	.775	59	.441
	B	.30189	.86791	2.532	52	.014
3	A	-.06667	1.23325	-.419	59	.677
	B	-.26415	1.02189	-1.882	52	.065

The pre-questionnaire variations can be partly explained by the perceptions of students based on their past experience, whereas, the post-questionnaire variations could be explained by a combination of the students' perception and their experiences during the course of the experiment. Table 4-22 shows that both groups had low expectations of their teacher's ability to motivate them, and in the case of group B, as shown in the second row of Table 4-22, this increase is significantly significant ($p=0.014<0.05$). Question 3 dealt with motivation from peers. Both groups gave moderate evaluations with no significant differences between pre- and post-tests.

4.4.9 Relationship between Psychological Factors and Peer Assessment Performance

The relationships between psychological factors as measured in terms of perceptions and attitudes and peer assessment performance are examined using the ANCOVA analysis of the preQ, postQ and the preQ-postQ means as covariates of the output of feedback quality and writing quality. The preQ and postQ means, as explained in previous sections, show changes in students' perceptions and attitudes after the experiment. All of the above factors were analysed using ANCOVA, but due to the large volume of data, only the statistically significant results are shown below in Table 4-23.

Table 4-23 ANCOVA psychological factors and peer assessment performance

Group	source	Dependent variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
A	PreQ1	Feedback1	10.819	1	10.819	6.884	.012	.144
	PreQ10	Writing2	4.843	1	4.843	5.350	.026	.115
	PostQ9	Feedback2	5.034	1	5.034	4.102	.049	.091
		Writing2	7.168	1	7.168	7.105	.011	.148
	Pre_Post2	Feedback2	8.890	1	8.890	6.867	.012	.119
	Pre_Post13	Writing2	5.987	1	5.987	5.656	.021	.100
B	PreQ12	Feedback2	4.643	1	4.643	16.623	.000	.304
	PreQ13	Feedback2	2.658	1	2.658	9.517	.004	.200
	PostQ1	Writing2	5.949	1	5.949	4.684	.038	.124
	PostQ4	Feedback1	1.206	1	1.206	5.581	.024	.145
	PostQ5	Feedback2	2.298	1	2.298	6.754	.014	.170
	PostQ8	Feedback2	1.407	1	1.407	4.135	.050	.111

	PostQ10	Feedback2	1.490	1	1.490	4.379	.044	.117
	PostQ11	Writing2	9.952	1	9.952	7.837	.008	.192

In group A, the influence of students' perceptions of and attitudes towards their peer assessment on pre-Q1, pre-Q10, and post-Q9 is significant, and the influence of changed perceptions and attitudes between tasks shows significant results for Q2 and Q13. However, previous *t* paired sample tests showed that the changes of perception between tasks were not themselves statistically significant (see Table 4-22 for Q2; Table 4-18 for Q13), and therefore the influence of Q2 and Q13 is not taken into consideration. In group B, the influence of perceptions and attitudes on peer assessment is significant for pre-Q12 and pre-Q13 post Q1, Q4, Q5, Q8, Q10, and Q11. Discussions of these results are given below in section 5.2.

4.5 Analysis of Classroom Interaction

Classroom interaction is at the heart of learning. In this section, classroom interaction among the teachers and the students are analysed to understand the underlying factors that impact on the effectiveness of the training and the limitations on students' learning.

Bernstein's (1996) concepts of classification and framing were used in this analytic with measurements derived from Morais (2002) (see section 3.7.5). These analysis techniques enabled the researcher to assess interaction between teachers and students moment by moment in a specific context. In this case, the classroom interaction was observed with regard to its contextual mechanics, such as the use of classification and framing. The observation records and transcripts provide evidence of the teacher's use of different techniques.

From the results of the analysis, a detailed picture of the interactions that took place during the peer assessment training could be assembled.

4.5.1 Observing Two Discourses

As in similar previous research (i.e. Min, 2005), there were two main tasks in the peer feedback training, where two different discourses were applied. The first discourse was used to introduce the criteria through modelling, in which the teacher introduced the four-step procedures and explained the implementation techniques. The second discourse included the negotiating criteria, in which the teacher offered 'substantive' and 'procedural' assistance. During substantive assistance the teacher worked with

students in modifying the comments that failed to follow the steps, and in subsequently responding to the revisions. Procedural assistance was designed to support students with questions regarding writing construction and then alerting them with oral prompts to the next step in the procedure.

In the following sections, excerpts from these two types of discourses are presented and analysed to recapture the key elements of the training.

4.5.2 Classification and Framing in the Discourses


4.5.2.1 Definitions and Notions

Definitions of framing and classification by Bernstein (1996) are included in section 2.4.2; however, for clarity, the following tables summarise these notions and give explanations of how they are used in the analysis in this section. The notation system is adapted from Morais (2002).

Table 4-24 Definitions of framing and strengths

Case of framing			Description of interactive relationships
Regulative	Weak ↓	RF--	Students' participation and intervention are not controlled by the teacher. The emphasis is placed on the highest possible degree of students' participation in the interaction process. In this case, the students are asked to solve open-type problems or to work independently. This requires cognitive or social abilities that presuppose a high degree of autonomy on the part of the student.
		RF-	Students' participation and intervention are not controlled entirely by the teacher. The emphasis is placed on the teacher's administrative role and a certain degree of students' participation in the interaction process.
	Strong	RF+	Students' participation and intervention are required in the act of teaching and/or the emphasis is placed on the teacher's administrative role.
		RF++	The statements emphasis the teacher's guiding role. This illustrates a passive stance on the part of the students in the interactive process.
Instructive	Weak ↓	IF--	The teacher accepts students' answers with a neutral attitude.
		IF-	The teacher or students point out what is incorrect but do not correct the error.
	Strong	IF+	The teacher or students point out what is correct / incorrect in a general way.
		IF++	The teacher or students point out what is correct / incorrect in a clear and detailed way.

Table 4-25 Definitions of classification and strengths

Case of classification		Description of the boundary between concepts/knowledge
Weak  Strong	C+	The knowledge of English argumentative writing is imparted through a strong argumentative concept and writing expression. The peer feedback steps are delivered in a specific order and with specific language.
	C-	The knowledge of English argumentative writing is imparted through argumentative concepts along with writing expression. The peer feedback steps are delivered in preferred order and language with problematic argumentative writing examples.

4.5.2.2 Modelling Discourse

In modelling discourse, the boundaries of classification have been presented by the introduction of the step concepts with an example from argumentative writing. During this discourse, the teacher has introduced the procedures and standards for producing peer feedback with examples that address a problem in the students' writing. In this discourse, the classification of writing instruction is strong while the classification of step training is relatively weak, which means that the teacher had to mark a clear boundary between writing and peer assessment instruction without paying too much attention to the step training. In the following excerpt, the instructive framing introduces 'The criteria and techniques for producing global aspects of peer feedback'.

Strong instructive framing (see Table 4-24) was observed when the teacher mentioned the goal for her instruction from turn 1 to 3 (IF++). In the following turns, she distributed the content and criteria in the form of an example of peer feedback. From turn 7, she commenced the explanation and instruction, using explanatory and instructional language; so that explanatory language includes 'bi fang suo', (for example) in turns 8 and 17, 'huan ju hua suo' (in other words) in turns 9 and 14 and 'ru guo' (if) in turns 14 and 19. This type of language was used by the teacher to offer explanations of the concepts, ideas, or information. Instructional language includes 'di yi ge' (first) in turn 11, 'di er ge' (second) in turn 14, 'jie zhe' (next) in turn 18. This form of language was used by the teacher to move on and accomplish a transition to the next topic. Finally, there is the use of confirmation checks such as 'Okay?' in turns 12 and 20.

Throughout all these turns, the teacher plays the role of ‘teacher’. The power of the teacher is strong; hence the framing is continuously strong.

Excerpt 1

- 1 T Zhe ge shi nimen yao zuo de san ge bu zo, nimen yao du nimen peng you de do:ng xi >jiu shi yao zuo zhe ge dong xi< Dan shi women xia, nimen yao zuo de Hao
These are the three steps you will later use for your friends in peer feedback activity. You will do this later. Fine.
- 2 T Women lai yong zhege tong xue xiede fanli lai xie, lai gau xu ta juei de na bian bu tuo=
Let us see this example written by previous students and use this to say what is not good
- 3 S4 Bu tuo
Not good
- 4 (.2)
- 5 T >ji ben de bu zuo< ni xien du wan zi ho xien xie chu. Xien xie xia lai ni juei de na ge bu fen ni juei de BU CHA DAN eh :: xie chu yao gai de di fan Okay ::Ni zhe ge shi yao bai da shenme ne ? okay >dao di zhe ge shi yao bai da shenme ne< ((finding files))
The first step, you read the text and first write down the part you think is not appropriate. Write down what should be revised, okay, what do you want to say? Okay, what do you want to say?
- 6 (.3)
- 7 T Hao di yi ge, ni yao xian xie chu yao gai de di fang. Ranho yao ba ba gai de di fang xie chu lai ni ke yi yong wenti de fang si wen duei fang li ching ni bu dong huoshi ni juei de : bu tai tuo dan de di fan
Fine, the first step you should write down is the part that needs improvement. Then, you can ask your friends to clarify the confusion or inappropriate argument.
- 8 T bifangsuo ni chao le yi ge ju zi ‘she also concern economical pressure problemes for her family, especially she plans to have a baby. The coasts, the cost must be a lot in the future’a ni suo qingwen ni zhe ge luo ji yao biao da ↓shenme(.)
For example, you wrote down ‘she also concern economical pressure problemes for her family, especially she plans to have a baby. The costs, the cost must be a lot in the future’ Then you asked what is the logic here?
- 9 T Huan ju hua suo ni bu tong yi ta zhe yang zi lai zuo jie shi ke ne bu go qing chu zhi le de. Ni jiao ta jing yi bu qu li qing.
In other words, why would the economical pressure for her family be a concern to sign up for the disagreement? You ask her to clarify the confusion.
- 10 T OK ni yong wen wenti de fang shi.((reading from the PPT)) ↑Ni zhi dao ma, yong wen wenti de fang si shi bi jiao keqi wei wang de fang shi ni liao jie ba. NI shi zai xie shenme gopidongxi, zhe yang zi jiu hen rude ho duibudui na women zai wen wen ti de shi ho jiu shi qing qiu dui fan lai liqing ni kan bu tai dong huoshi bu tou dan de di fan(.)
OK use clarify questions. You know what? Clarify questions is a better way to ask for clarification. If you say what are you writing about, I don’t

- understand at all. It is rude, it's better to use questions to clarify the confusion*
- 11 T OK ↑di yi ge bu zo xien xie chu yao gai de di fan.
OK the first step, write down the part that needs revision
- 12 T OKAY ?
- 13 T Na di er ge shi xie chu wenti dien. Weishenme nijuei de zhen li xu yao xiu gai. Er qie ni yao jie shi zhe ge wen ti dien de zong yao xing.
Then the second step, write down the reason why this is a problem that needs revision.
- 14 T HUAN JU HUA SUO RU: GUO: ruen dien bu go xiang er you li ke ne hui ying xiang wen zang de suo fu li ZHE YANG jiu shi yao jie shi wei shenme ni jei de zhe ge di fan hai bu go hon?().
In other words, if the logic is not clear here, the argument would be less persuasive. Explain why you think this part is not good enough.
- 15 T HAO
Fine
- 16 T NA
Then
- 17 T bi fang suo wo juei de zhe ge luo ji de wen ti ne zai yu tai de jing ji kao liang yu ta tong yi zhe ge zhen che de guan xi bu qing zu zhe yang zi hui rang zhe ge run xu bu go you li JIU SHI li zi ni ke yi zheme yang jie shi WO JUEI de ne zhe ge di fan ying gai xu gai ta ying gai gen zi jei de lai jie shi suo er jing ji de ya li gen ta yao bu yao zi chi zhe ge zhen che zhi jen de guan xi jiushi mei you jie shi qing zu duzhe ke ne hui kan bu dong huo shi ni de jiu bu go shuo fu ren zhi lei de
For example, I think the point of economic concern and her act to sign up for the disagreement is not strongly related to make a strong argument. I think it is better to be clear about the link between the economic pressure she got and how this links to her action. This part is not clearly explained. If this is not clearly explained, the argument is weak and difficult to understand.
- 18 T HAO jie zhe ni yao jiang ni ju ti de jien yi ru guo ni shi nage WRITER de hua RUGUO ni huan zai ta de jiao du ni hui zhenme gai().
Now you need to give a concrete suggestion to the writer. If you are the writer what will you change?
- 19 T RUGRO ni suo jessica bu tong yi zhe ge zhen che shi ying wei jing ji de ying su, ni ke yi zhe yang xie chu lai >zai dan xia jing ji xiao tiao de shi: ho: ru guo ta xiang yao chun qien ge xiao< hai yong ta xien zai bi xu bao quen ta de gong zuo er fu chung gong si ding xia de zhen che ke yi rang ta bi jiao bu hui diu diao zhege gong zuo ji hui. NA zhe ge jiu gen chie ti jing ji de ya li gen ta yao zhi chi zhe ge gong si zhenche zhi jien de↓ guan xi.
If you say Jessica disagrees with the policy because of her economical pressure. You can revise in this way: during economic down turn, it is better for her to sign up for the agreement to keep her job. To link the ideas in this way, it would stress the point you tried to make about economic pressure and how it related to the act and the disagreement.
- 20 T OK?
- 21 T HAO
Fine

The regulative framing is also strong (RF++) in this context when the teacher has dominated the interaction with corrective evaluations. In turns 1 and 2, the rule was established that the teacher was going to introduce and lecture about the standards for students' later peer group work.

There were also many occasions when transition words were used before the teacher gave the explanation of each step, such as 'hao'(OK) in turns 7, 15, 18 and 21, 'na' (so) in turn 13 and 16 and 'OK' (in turns 10 and 11). The frequent and intensive use of the transition words in excerpt 1 shows the teacher's intention to distribute the concepts of criteria to the students in a quick and concise manner without giving long and detailed statements of explanation. All these statements/claims placed an emphasis which underlined the 'teacher's' authority. By doing that, the teacher is setting up a clear boundary of the content to be delivered.

On the other hand, and as expected, a passive stance on the part of the students was observed during the interaction. There are only two short turns taken by S4 taken at the beginning of the excerpt which are then followed by the teacher's extended turns. The lack of student turns shows the weaker power of the student during the interaction and reinforces the power that the teacher retains during the interaction.

4.5.2.3 Negotiation Discourse

The negotiation discourse is expected to be different from the modelling discourse. The teacher used 'steps' as a framework to elicit students' responses and their understanding of the issues in the discourse. To allow the interaction to flow, the peer feedback steps are delivered in the participants' preferred order and language. The teacher invited students' comments in order to revisit and revise the content of the identify step and explain (e.g. Excerpt 2, turn 13). In contrast to the modelling discourse, the boundary of classification between writing instruction and steps usage is weak.

In a departure from the modelling discourse, two types of framing were observed:

Type 1: strong instructional framing and strong regulative framing (IF+ and RF+); for example, excerpt 2 and 3.

Type 2: strong or very strong instructional framing but weak regulative framing (IF+/++ and RF/--); for examples excerpts 4 and 5.

Type 1

In excerpt 2, the regulative framing is moderately strong (RF+) when the teacher summarized the discussion quickly after attending to class management due to noise. She then asked S5 two questions, for which she had answers in mind. These questions were asked in the previous turn and again here as a class management strategy, during which her aim was to re-engage the whole class. The two questions were originally used to create debate between the S12s group and their opposite group. S12 adhered to the rules and took the next turn (15). As in the debate mode, she used ‘we’ for her group and ‘they’ for the opposite group. In this excerpt, the regulative framing was strong (RF++). The teacher did not invite the opposite group’s opinion, but instead used turn 16 to correct the errors in the statement and request further clarification of certain details. Under this strong regulative framing, the interactive rules changed at this point to an iInitiate-response-evaluation pattern. The statements emphasise the teacher’s guiding role and highlight a passive stance from the students in the interactive process. S5, who thought she was under assess, was eager to clarify the answer she provided. In turns 18 and 20, the teacher had to highlight the problem with a direct repair to correct S5’s answer.

At the start of this excerpt, the instructive framing is moderately weak (IF-), as the two questions in turn 13 are ambiguous. The first question asked about a reason and the second question about an explanation of an action. The teacher’s real purpose was not to invite brainstorming. Instead, she intended S5 to repeat her response to bridge for the next turn. As a result, the instructive framing turned to a strong level in turn 16 (IF++) in which the teacher pointed out what was correct/incorrect in a clear and detailed way. From then, the instructive framing became consistently strong to help S5 produce correct answers (turns 18 and 20).

Excerpt 2

- 13 T U::h qitade zu you biedeyijien shibushi. xien zuyi yi ixia zhe yizu tamen zai zichu ling wai yi ge fangxiang de ↓wentī (.) qizong yi zu renwei↓(.) yuanlai zhejia gongxi bingmeiyou MAI shiwu de jing yen (.) SOUYI ta:men YINGGAIGEN lia:o jie:↓ MESAFOODTade shi:pi:ng xi:ngxiao: che:lue: gen(.) yi:ngyu:n de fa:ng fa:: CAI ke yi(.) ZA:NGK:ONG weilai >jiushi< BINGGO(.) shibushi dui(.) zhe duiyu zhe u::h >weilai juishi mai zhe jia gongsi shibushi ZHENGzhen de huoli<(.) NA TAMEN RENWEI ZONGDIEN BUZAI YU(.5) Om Omni ZHE JIA gongsi: YOU:MEI:YOU: XIAO:SHO:shipingde JINGYEN(.) ZONGDIEN shi yinggai FANGZAI SUI:.(.) YAO:.(.) Mai:.(.) zhe yi ge(.2) ↓shiping (.)

- >tamen de kequeng shi sheme zongdien ying gai fan zai zhe shangmien<
A:: nimen butongyi >tamende suo fa(.5) duibudui< nimen butongyi (.)
weishenme nimen buner jiesho↓ >tamen geinimende jienyi?<
*other groups have comments too? Please pay attention to this group ...
they are trying to point out a different point of view. One of the group
considered that Omni company doesn't have the experience to sell snacks
so they should know about the marketing strategies and management so
they can gain control on the future of Mesa food and evaluate the benefits
to merge the company. The other group thinks the focus is not about
whether Omni company has experience selling snacks or not, instead the
focus should be about who the target customers of AA salsa are and who
is going to buy this food. (Turning to S8) And you do not agree with the
other group's argument. Right? Why not? Why can't you accept their
suggestions?*
- 14 (.)
15 S5 yingwei tamen gangan jiangde jiushi women you xiezai la:: suo kequeng
jiushi cheluede qizong zi yi a:
*because we have already written down what they suggested. Target
customers would be part of marketing strategies.*
- 16 T Ni jueide kequeng shi chelue[de =
you think target customer is part of strategy
- 17 S5 [>Jiushishuo< tamen bushi shuo
ZONGDIEN yinggai fan zai kequen ke shi kequen jiushi women yijing
you xiedao((point to the black board))[° zhe yi kuai=
*That is, the other group said the point should
be placed on the target customer but we have written about this point of
view, the target customers*
- 18 T [KESHI>nimen yi kaishi jiangde shi< MEI YOU
↓[JI:NGYEN=
but at first you talk about no selling experience
- 19 S5 [NA(.).shi ruguo la::dui jashe=
Just,.it is a 'if' yes an assumption
- 20 T MEI YOU ↓[JI:NGYEN(.) A:: gen chelue youshenme guanxi?
No selling experience?How is this related to marketing strategies?
- 21 S5 [Jiushiwomen xiyao=((looking at S11))
It just means that we need to
- 22 S11 [a: womende zo::ngdie::nshi::ZO::NG DIE::N bushi suo::
women maiyou jingyen zongdien shi: suo:: women yao zhidao mesa food
de xing xiao shuo fa gen chelueNA gen womende shichang cha henduo de
hua jiu meiyou banfa qiu caozuo a: shi zhe yang
*what we try to say the point is, the point is not about no selling
experience, the point is the need to know Mesa Foods marketing
strategies and that is very important information to see if our target
customer can accept that (food), if there is a big difference, then we
probably cannot buy this company.*
- 23 T (big Smile)Tamen de ru;ndie;n ni;me;n do;ngma?((turning to group
14))(.) Haiyou meiyou shenmeyangde comment?
Do you understand their reasoning? Do you have any further comments?

A similar pattern of framing is also found in excerpt 3. The instructive framing is weak (IF-) when the teacher invited students to comment on the previous students' turns (RF+). Unlike in excerpt 2, the instructive framing became moderately strong in turns 72 and 74 (IF+). In both these turns, the teacher pointed out what was incorrect in a general way. In turn 74, the teacher concentrated on the strong classification of writing instruction, and did not correct S2's answer for the way she constructed peer feedback, instead correcting the content of the step. This moderately strong instructive framing and classification on writing instruction gives explicit criteria for evaluation, but denies the students the opportunity of learning the legitimate text on their own. The students were confused and at the completion of the turns, failed or refused to try again and produced answers when the regulative framing, in turns 77 and 78, was strong (RF++).

Excerpt 3

- 63 T OK hao diyizhu(0.2) °zai na li?° zhendueita:mende(.)zhegelundien ni:men
you:: shenmede(.) nirenwei you shenmeya:ng defang(.) xi::yao xiu zhen↓ huo
nabien bu tuo,[heishi shuo=
*OK, fine, Group 1, according to what they say.... can you ... do you want to
revise anything ... do you think there is anything not good ... or you want to
correct ... comment on?*
- 64 S2 [juishi(.)
There is er ...
- 65 T =shen[me?
What?
- 66 S2 [wojueide tadeliyouhaoxiang°buguo(.)dui°>yingwei(.)suiran shuo
women zhidao tamen xiang yao zhi dao zhen baifenzhiershide touziren
huoliduoduo(.)NA ruguoshuo na jiarushuo nage mesa food gen ta men suo
xien zai you duo shao ren you duo shao jing er bu jiu gen ta suo yao gen ta
hebing<na zhe yang bu jiu you dien ((hands waving)) mao: mao dun
ma?((touching hair))
*I feel that ... just the reason ... the reason seems not good enough ... although
we know they try to argue with the 20 per cent gain and the number of
shareholders ... what if Mesa food give them the information they need, will
they (the company) really work on the investment just because of that
information ...it seems ... just that a little bit con ... contradictory*
- 67 (.1)
- 68 T A:h ah ni yao zhen me gai zhe ge li you?
Ah. how will you revise this?
- 69 (.5)
- 70 S2 .hhh. juishi(0.5)wo juide ta ke yi jieng yi na ge (.)MESA food na ge go:ng
si:, juishi baochulai yao zhenme °fenpeigeigudong↓° juishi bbiaoda bijiao
ming chei (zheyang)°
*I think they can suggest Mesa Foods to provide the information about the
distribution of the 20 percent gains to shareholders and (inaudible)*
- 71 T °O:h, okay:° SUOYI yao ti gong zi xueng(.) guan yu: MESA Mesa gong xiu

- yao ti gong guan yu ta zai °zhe ge°GUDONJIEGO↓ yao zhen me >qui fen
pei na ge baifenzhi ershi de huo li<(.1)
*so ... they should suggest Mesa foods to provide the information about the
shareholders and how the 20 per cent is shared by the stakeholders*
- 72 T uh(.)zhe yangzi gen niMA:IBUMA:I zhe jia gong si you guan xi ma?
...uh...so is this related to the issue of buying Mesa Foods?
- 73 S2 Uh(>)zhe shi women< gei ta de jien yi((pointing to S1))ta yi jingshi wang
zhebian fanzan(.)DANRAN wo shi juide shuo ta suo zhebian ((pointing the
memo article on her table))shi mai you shuo yao feng gei sui hai shi yao fen
duo shao gei gudong ranhou women ganggang na ge shi gei ta de jien yi. (.3)
((uncertain laugh))
*Uh .. this is ... this is the suggestions we gave them if they want to keep this
argument and develop the argument... of course we think thisuh...
because from what we read in the passage there is no such information
about shareholders or regarding how the shares are distributed ...what I
say earlier is the suggestion only to help them to revise this
argument...(laugh, not sure)*
- 74 T °Ok° (.)
Ok.
- 75 T NA ni xien zai YAO xie yi ge fan dui gong si (.) bu ying gai xie bu ying gai
mai mesa zheyi zheyi jia gong si(.) NA ni:: yao zhen me yang qui shei fu
gen qui zhe ge gu dong li yi zhen me fen pai de tiao jian NI JUEI de gen
maibumai you mei you guang xi ?
*Then, now you are writing a letter to disagree with the investment. So how
are you going to argue with the reasoning with the shareholder things ... do
you think this reasoning is related to buying Mesa Foods or not ?*
- 76 S2 ((talking to her peer quickly(.5))
- 77 T WEI SHEN me BU YAO mai?
Why not buy Mesa foods?
- 78 S2 ((talking to her peer quickly)(.10)
- 79 T Okay(.) ((turning away and face all class))hai you mai you qi ta zu(.) you
mai you bai de YI JIEN ? (.) guang yu ta men zheyi zong runzhen ni juei:
de(.)you mai you guang xi? Huozhe: shi get amen shenme yang de jien yi?
*Okay what about other students? What do you think about this argument?
Do you think the reasoning is related and convincing? Or anyone want to
comment on this?*

Type 2

In excerpt 4 turn 121, the regulative framing is moderately strong (RF+) when the teacher summarized the discussion and invited students' comments. She organized a debate between S5 and S6. In this, students' participation and intervention is still controlled by the teacher. In turn 123, the teacher decreased the regulative strength (RF-) where she chose not to select individual students to take the next turn but instead created a potential debating topic for students to participate in. In turn 125, the teacher gave a prolonged 'okay:::', indicating her hesitation about further controlled turns. This hesitation signalled a possible turn, with the chance that S4 could join the discussion

without reference to the teachers' admission in turn 126 (RF-). Then, the teacher's overlapping turns suggests her control over the turn taking in 127(RF+). In turn 130, S4 was granted a turn (RF+) to speak about his argument. Then, in turn 131, the teacher ended the discussion by rephrasing the student's argument herself.

Excerpt 4

- 121 T Tasuo women bukeyi(.)ta fandui yaomai mesa(.) zhejia gongsi YINGWEI women BUZHIDAO zhe baifenziershi de >huoli<SUO daibiao de shi duoshao ↓qien huoxi zhege huolide jinger bingbuzide women gonsi qiu touzi: >JIU tade yi si kener women gaobuhao< haiyao touzi genduo RAN ho huoli ziyou yidiendien SUO yi ta ↓FANDUI. A: ni jiede genwomen zhebiende(.) yi ti youmeiyou XIA::NGGUA::N?huoshi youmeiyou shenme WENTI?
she said that she disagreed with the merging because she doesn't know what the 20 per cent means in money and maybe the amount of money is not worth the investment. So what she meant could be the money we put in may be more than what the merging can bring us. So she doesn't agree with the investment. Do you think the argument related to the issue we brought up here? Any there hidden problems here?
- 122 S5 Wojiede shi ok. DANSHI yingwei tanage shi small region(.) ruguo women shiyao sell in big region dehua wo jiede nage baifenzhiershi jiuhui gouda le.
I think it is OK. But the 20 per cent profit gain is in small region so the money can be small but if it was sold is in a big region. The money can be bigger.
- 123 T ((waving to get S6's group's attention)) S6's group(.) tageinidejienyi suo u::h tamen baifenziershi shi XIAOQUYU de huoli okay: a: DANSHI women xienzai yao ba ta zuanhuang chen DAQUYU a jiu shi jiaru xiang ta jiangde huiyou baifenziershi de huoli kuo da dao daquyu dehua na nage jinger ye hui kuo da.(2)suoyi percentage yinggai bushi nide arugment de (.) de dien
They say the 20 per cent gain is in the small region, but when it is in a big region the money can be huge. So percentage is not what you should argue for.
- 124 S6 KE::shi>zai xiaoquyu zhenjia↓ baifenziershi ERQIE shi QIUNIEN zhenjia baifenziershi TABUSHI yucheni mingnien zai DAQUYU yao zhenjia baifenziershi<NA shi LIANGHUIISHI
but it says that it increases the sale to 20 per cent and it was last year but it doesn't mean and cannot be a prediction for the 20 per cent growth to be in the big region or in the future.
- 125 T Okay:::
- 126 S4 SUOYI [wo caihui xie zhege wo wo wo=
so that's what I talk about ... why I would talk about that (pointing to the statements written on the blackboard) in my argument. .
- 127 T [sou yi nimen zhezu de idea°=
So your group's idea
- 128 Ss [((Ss laugh))
- 129 T Na yi zhu? Zhe yi zu. Haohao.)(smile)) gankuai jiezhexiachu.°nimende argument shishenme? qwuai jiang° ((background noise:Ss laughs))
which group? (smile) fine, fine, this group, let's continue the topic, what's your argument?quickly
- 130 S4 Jiu JIU suo yi wo cai xie NI YAO bingo yijia gongsi wo yiding xiangyao zhidaosuo WEILAide nage zui duoshouqien pasuduoshao WO BIXUYAO zhidao WENGBUWENDING(.) dui: a::
So that's I what I wrote ... that's why I say ... when you need to merge a

company, you need to know about the potential of this company, for example, the margin of growth, financial problems and so on. We need to know about these information

131 T Ok hao, suoyi S4 de zu Tamen renwei(.) GENMINGCHEI: de jiang guochidehuoli bu daibiao weilaidehuoli wo women xienzai gen YINGAI pingu ta weilai you duodade huoli potential okay::

OK fine, so S4's group think, in fact, the profit does not mean future potential. More importantly, we should evaluate its future potential

The instructive framing in this excerpt started weakly (IF-) when the teacher summarized the discussion and tried to invite comments in turn 121. In turn 123, the strength of framing increased when the teacher reshaped the topic (IF+) to tackle the problem. In turn 125, the teacher gave a neutral evaluation of S6's response, suggesting a weak form of framing (IF--). In turn 126, S4 provided a clear and detailed explanation of the problems identified which were further elaborated on with additional detail that had been lacking in the previous teacher and students' turns. As a result of this, the framing became strong (IF+). The strength continued in turn 131, and the teacher repeated S4's points of view with 'more importantly,...', an expression showing the teacher's positive evaluation.

Excerpt 5 has a similar combination of framing to that in excerpt 4. The instructional framing is moderately strong (+) when the teacher suggested that students identify problems with explanations and invited their comments. The framing required S9 to respond to the question (RF+). It was noticed that S9 slipped away (talking to her peers to avoid this turn taking) because she was not confident to explore the answers with the teacher. In turn 169, the regulative framing becomes moderately weak (RF-) when the teacher waited patiently, keeping the discussion open. Then, the regulative framing in this context becomes weak (RF--) when the teacher waited for students to read through the table and get ready in turn 170. In turn 171, S10 raised a hand to get the teacher's attention. Although by doing so he requested permission to join the interaction, he decided for himself when he was ready to participate. So the regulative framing in this context remains weak (RF--). In turns 174-179, the framing remains weak when S10 joined in the conversation without requesting the teacher's permission with three overlapping turns (turn 175,177,179).

In turn 169, when detailed information was provided, the instructional framing becomes strong (IF++). In this case, the teacher provided clear and specific information and

criteria for students who wished to take the next turns. In turn 173-179, S10 gave answers and responded to the teacher's confirmation, which corresponds to turn 169.

Excerpt 5

- 166 T Hao na di erge reason YO::U banfa qiu suofu ni dao BUYAO(.) jiushi jienyi
gongsi buyao mai?
*how about the SECOND reasoning? Do you think the SECOND reasoning is
GOOD enough to convince you? (0.5) to give up the investment.*
- 167 (.2)
- 168 S9 °youma?((talking to her peer))°
is there?
- 169 T TA suo chanpingde tesebu:zhi::dao suoyi gongsi bu:yinggai mai MESA food
(.4) zhe shi information ((pointing to ppt slides))
*he said it is not good to merge the company because there is no information
about the qualities of the product here are the information ((pointing to ppt
slides))*
- 170 (.10)(Most of Ss are watching and reading the visual aids)
- 171 S10 ((Raised hand))
- 172 T Hao.
Fine.
- 173 S10 Uh::wohuigui naju(.)>wo juei de hai shi yao hui gui dao xiaofei zhe de xu
qiu ER:BU:SHI: suo qiu liaojie changing de tese shishenme jiu keyi
uh::<(0.5) jiaru dozhidao jiu denyu suo yao qiu biao shi wo yao qiu bingo
zhejia gongsi °le ma° ((sit up and lean forward))?
*Um I think we should go back to the key ideas, it is the target customer
should be the focus here not the feature of the product. suppose we know all
the features of the product, does it mean that we are going to buy this
company?*
- 174 T SUOYI NI renwei(.) ni renwei CHANGPING de tese [buyao=
*So you think the qualities of products are not a concern here. It is more
important to understand the target customer and why they will buy this
product?*
- 175 S10 [bing bushi
It is not
- 176 T =binbushi zhege [zongdien
Is not important here
- 177 S10 [dui
correct
- 178 T SUI yaomaizhege chan ping cai shi [zongdien=
Who is the customer should be the focus
- 179 S10 [dui
correct

From the observation above, the pattern of framing of the negotiation discourse can be summarized as follows:

Figure 4-1 Framing in the negotiation discourse

T's turn 1	S's turn 1	T's turn 2
IF- RF+	→ Examined under strong classification	↗ IF+RF+(type 1) ↘ IF+/++RF/--(type 2)

Figure 4-1 describes the framing that started at the beginning of a new topic. T's turn 1, is followed by the students' response in S's turn 1. At that moment the teacher has made a judgement about the answers with the criteria from strong framing. T's turn 2 shows how the teacher reacted to the students' errors. This influenced the subsequent turn-taking in both content and motivation.

In the two excerpts 4 and 5, the framing for a new topic started with the similar condition (RF+IF-). After several turns, it evolved into two different forms of framing. The first type of framing appeared when the teacher made the criteria of evaluation explicit (IF++). When the students were requested to take turns in the interaction, the level of regulative discourse is "imperative/positional" (Morais, 2002) (RF+/++). Occasionally, when the teacher invoked a moderately strong instructional framing (IF+) in T's turn 2, the risk was that students' were not allowed to explore the answers but to respond to the request by simply guessing (Leung and Mohan, 2004). The imperative/positional power relations indicate the importance of the teacher maintaining strong instructional framing in this situation.

The second type of framing appeared when the teacher also made the criteria of evaluation explicit in T's turn 2 (IF+/++), but in this instance the teacher used personal control (RF/--). The students are accorded the means to discuss the mark/criterion and the responsibility for maintaining strong framing is in their hands. This allowed students to recognize the criterion and apply the criterion to their peer assessment. The personal power relations demonstrate the responsibility for empowering students to develop strong instructional framing which can lead to a higher level of learning engagement than in other types of framing.

4.5.3 Teacher's Role and Techniques

The flexibility of the teacher's role change is crucial to the structure of the students' learning (see section 2.4.3). The combination of classification and framing reflects the

change in the interactions which also assist students' learning. In this section, based on the classification and framing previously observed, the teacher's techniques for more effectively managing their change in role are examined to further test Pryor and Crossouard's (2008) theory of the role of formative assessment in students' learning.

4.5.3.1 Teacher's Role Examined in the Transcript for Type 1 Framing

Type 1 framing is strong regulative and instructional framing. In the following sections, excerpts 2 and 3 as examples of such framing are further examined for their interactive features.

In excerpt 2, the teacher started a debate with a summary and a referential question. Then, the teacher constructed a strong regulative and instructional framing scenario, introducing strong classification when she found students' errors. When S5 gave an incorrect answer in turn 15 the teacher (turn 16) repeated the problematic statement. In turns 19 and 20, the teacher used a direct repair to point out directly the problematic statement which required S2 to revise her answer. These skills indicated the positional roles of the teacher and student in the relationship - in this case, in assessor-learner roles.

In excerpt 3, the combination of framing is very similar to that in that in excerpt 2 but its instructional framing is weaker and an examination of interactive features is necessary. After S2 has produced her response in turn 67, the teacher asked a question in turn 68 for a suggestion to solve the identified problem. It is a referential question, since the teacher expected S2 to provide an answer related to her previous turn. The teacher did not give any guidance on how the suggestion should be constructed, but instead provided a neutral evaluation in turn 71 with a summary of S2's response. In turn 72, a clarification request was made of S2 to indicate a problem in her response. In turn 73, S2 tried to explain that she was aware of the problem but the students decided, nevertheless, to make improvements based on the problematic argument. However, it could be proposed that the best method to deal with the problematic argument was to remove it. The teacher should have intervened to stop the argument from developing further. At that point, however, the teacher was in a relatively weak position and in turn 74 another neutral evaluation was provided followed by another clarification request in turn 75.

In this excerpt, the teacher was in a weak role throughout the interaction and was unable to apply any strong classifications to the interaction. She used the teacher's authority to guide the conversation but did not have the chance to provide instructions to correct the students' problematic arguments. The reasons why the teacher did not fully address the problem are not clear although this kind of omission was common. The consequences of this lack of instructional discourse were clear in that the students did not have the chance to rectify the problems during the duration of the discourse. In this excerpt, the interaction between the teacher and students had strong regulative framing, where the teacher dominated the direction of the discussion, but had weak instructional framing so that the teacher did not provide instructions to correct the students' mistakes. The teacher led the conversation using her authority; hence the teacher's role was as the 'educator' (or 'teacher'; the term educator is used here for clarity). The effect was that the interaction followed the teacher's will, but the students were left with an uncorrected and problematic argument.

4.5.3.2 Teacher's Role Examined in the Transcript for Type 2 Framing

The type 2 combination of framing is consists of moderately weak or weak regulative and strong instructional framing. The following section examines the interactive features in excerpts 4 and 5.

In excerpt 4, the teacher invited comments with critical evaluation, which she made available to students by using referential questions such as in turn 121, which was a genuine question to which the teacher did not know the answer. This confirmed her role as teacher during the topic. Then, clarification of meaning (in turn 123) allowed the students to reflect on what had been said and to anticipate what might happen, enabling them to take part in the following turns when ready. Then, S6 joined in. In turn 125, the teacher gave a neutral evaluation of S6's response, thereby reducing her role as assessor. At this moment, S10 had the chance to be evaluative and to voluntarily join in the discussion. When S4 started to talk in turn 126, the teacher interrupted but granted permission, in 129, for S4 to continue. S4 was ready to talk but not long after that the teacher provided a confirmation, demonstrating her role of assessor in confirming the answer. In turn 131, the teacher took the 'assessor' role. A turn-taking indicates that the teacher did not wish to pursue further discussion, so the role of 'teacher' was resumed and the authority of the teacher was re-established very quickly. She checked S4's response by offering a summary that was appropriate but did not offer further learning

scaffolding. This assessment result was announced with a high and flat voice as a signal to close down the discussion and to move on, showing the teacher's control of the topic. As a result, the discussion was effectively terminated and no more student activity could be pursued. This cut-off showed the immediate impact on the interaction of power shifting from the students to the teacher. Excerpt 5 shows a similar combination of framing as that in excerpt 4. At first, a referential question was asked to invite comments from S9 (turn 166). When the teacher realized that S9 had encountered problems in turns 167 and 168, she responded to the problem with visual aids (turn 169). This response encouraged more students to take part in the interaction and the subsequent extended waiting time (turn 170) was crucial to allow the students the opportunity to think through the questions and answers. Up to this point, the teacher had acted as a knowledge expert. Then, the teacher checked that she had understood S10's response correctly with two clarification questions (turns 174 and 178), a strategy that allowed her to extend her evaluation of S10's response. The student, therefore, could produce the answers at his own pace.

From the above, the techniques used by the teacher and the roles played in the above excerpts are summarised in the following table.

Table 4-26 The techniques and teacher's role in 2 types of framing combination

Excerpts	Framing	Teacher's role	Teacher's techniques
Excerpt 2	IF+RF+	Teacher and assessor	Teacher's echo, direct repair
Excerpt 3	IF-RF+	Teacher	Neutral evaluation
Excerpt 4	IF+RF-	Knowledge expert	Reformulation
Excerpt 5	IF++RF- -	Knowledge expert	Task reformulation, wait time, clarification question

4.5.4 From Convergent to Divergent Assessment

Pryor and Crossouard (2008) proposed that formative peer assessment could be best achieved through the teacher's flexible role change. This role should allow the formative peer assessment to move from convergent to divergent assessment. For

example, the teacher should take the assessor's role at the convergent end of assessment but at the divergent end, the teacher should assist students' learning in the knowledge expert/learner role. This section examines how this kind of movement could be achieved in pedagogical discourse.

For this purpose, excerpts from both modelling and negotiation discourses were analysed to find evidence of role changes and the movement between assessment types within the interaction. Excerpts 2, 4 and 5 were all found to fit this purpose. Three students, known as S4, S5 and S10, participated in all 3 excerpts. The transcript was taken from 3 different time slots, namely turns 13-23, 121-131 and 166-179 in the same session.

For excerpt 1, the analysis result of framing (see section 4.5.2.1) was characterized by strong regulative and instructional framing (IF++ RF++), suggesting its position on the extreme convergent end of the assessment continuum where only the teacher's role was observed. Excerpts 2, 4, and 5, on the other hand, showed different aspects of divergent assessment.

The framing in excerpt 2 followed the direction of IF-RF+ to IF++RF++. The strong framing characterised the teacher's 'teacher' and 'assessor' roles in the interaction, and suggested that the assessment was still at the convergent end of formative assessment.

In excerpt 4, the framing followed IF- RF+ to IF- RF +, which suggested that assessment was at the middle point in the continuum, moving towards the divergent end of the formative assessment. The knowledge expert role of the teacher was observed but this was soon changed back to the role of 'teacher'.

In excerpt 5, the framing followed IF- RF+ to IF++RF--, which suggested that the divergent end of formative assessment had been reached. The knowledge expert role of the teacher was observed up until the end of the excerpt, which then anticipated the next role change (to that of learner) in excerpt 6 (see section 4.5.5).

As a result, the continuum of convergent and divergent assessment for step training within writing instruction can be summarised as follows:

Figure 4-2 Formative assessment for step training

	Continuum of formative assessment		
	convergent	←→	divergent
Example excerpts	Excerpt 1/2	Excerpt 4	Excerpt 5/6
Framing	IF++ RF++	IF+ RF-	IF++ RF--
Teacher's role	Teacher, assessor	Knowledge expert/teacher	Knowledge expert, learner
Learner's learning	Recognition	Passive realization	Active realization


The interactions shown in the excerpts show the movement from convergent to divergent assessment. Each excerpt represents one stage in the continuum. In this continuum from convergent to divergent, the framing changed from RF++ to RF--; meanwhile the instructional framing changed less than the regulative framing. As a result, the teacher's role changed from the single 'teacher' role to the 'teacher-assessor', the 'knowledge expert-teacher' and then the 'knowledge expert-learner' role.

From convergent (using modelling discourses) to divergent assessment (negotiation discourses), the students learned to 'read' the context in such a way that enabled the development of legitimate understanding and the production of legitimate text. This process of learning is summarized in Table 4-25.

By the end of excerpt 1, the students could develop a recognition rule which indicates the principles and criteria they need to follow when producing the steps. In addition to the principles for producing steps, at the same time students also needed to develop their understanding of the concept of argumentation.

In excerpts 2, 4 and 5, such development was shown in different degrees of recognition and realization. In excerpt 2, turn 17, S5's answer suggests that the recognition rule is still developing because she struggled with the relationship between concepts and argument so that her attempt to produce proper step identification and explanation failed.

Table 4-27 Convergent to divergent formative assessment for learning

<div style="text-align: center;">  </div>	Convergent	Students' feedback from excerpts	Acquisition
		(Excerpt 2 turn 17)S5: That is, the other group said the point should be placed on the target customer but we have written about this point of view, the target customers	Recognition: four steps
		(Excerpt 4 turn 130)S4: So that's what I wrote that's why I say ... when you need to merge a company, you need to know about the potential of this company, for example, the margin of growth, financial problems and so on. We need to know about this information.	Passive realization: identify problems with explanations critically.
	Divergent	(Excerpt 5 turn 173)S10: Um, I think we should go back to the key ideas, the target customer should be the focus here not the features of the product. suppose we know all the features of the product, does it mean that we are going to buy this company?	Active realization: Thinking about the problem in a specific way.

In excerpt 4, turn 130, the students were found to be well aware of the concepts mentioned in the pedagogical discourse. They demonstrated sufficient knowledge of the topic, and also that they were ready to provide answers to the questions. S4, for example, joined the discussion voluntarily. His answer suggests that he could 'read' the context clearly and gave an appropriate answer to the question even though this answer was not produced in the most effective manner and the teacher reformulated the answer in the following turn. At this point, the students had developed recognition rules but produced 'passive realization' (Bernstein, 1997), and were not able to produce the answer which followed the principles of identifying steps and explanation.

In excerpt 5, turn 173, S10 demonstrated his understanding of the issues in the argument in an appropriate way and which followed the correct usage for step identification and explanation. This characterises the students' learning stage in 'active realization' (Bernstein, 1997), at which point the students could make the meaning public in an appropriate way and without prompts.

4.5.5 Construction of the Student's Identity in Peer Assessment Training

The previous results cover the role change of the teacher, but it was more difficult to capture the exact identification changes among the students. This was due to the large class sizes and the difficulties this presented in mapping the ongoing changes in roles of

individual students. However, by analysing the following excerpt, some preliminary evidence of such changes can be found.

In excerpt 6, from a continuous transcript with excerpt 5, the teacher and student S10 continued the discussion from excerpt 5. In turn 180, the regulative framing was moderately strong when the teacher invited S10 in the next turn to produce the step suggested (RF+). In the same turn, the teacher then suggested her solution (IF+). This suggestion placed the teacher in a vulnerable situation by allowing the student to make a judgement of her ideas. The pause in turn 181 led to a revision of her own suggestion which incorporated S10's idea in the previous turn. The framing reached a strong level (IF++) when the teacher provided a detailed example for consideration. Again, the teacher put herself in the learner's role for evaluation. The teacher retained weak regulative framing (RF-) in this turn, waiting for S10's readiness to make a decision. In contrast to the previous excerpt, the student's knowledge (as demonstrated in the previous stages) and their understanding of the knowledge in legitimate text enabled S10 to evaluate and consider these 'learner's' suggestions carefully. In turn 183, the student used 'I think' to begin his answer. However, the student's evaluation of the suggestion was not presented with suitable reasoning. Despite this, the confidence level of S10 was high and he was able to arrive at his own evaluation of the teacher's suggestions. This may be evidence of the identity change of the student as a result of the interaction with the divergent assessment indicated in excerpt 5.

This analysis demonstrates how the interactions with different teacher roles and framing could shape the student's identity. The teacher's different roles and framing control could directly influence the student's response and confidence level. Although direct changes in student identity were not observed in the excerpts, such training could eventually invoke the student's identity during a feedback task. Hence it will help in the construction of the students' identity in peer assessment.

Excerpt 6

- 180 T Okay okay: suoyi ni jien yi TA yao zheme gai?(.1) jiu bao zhege shandiao?((pointing to sentence on the board))
 Okay okay, so, what do you suggest to revise his argument? Delete the second reason?((pointing to sentence on the board))
- 181 (.2)
- 182 T Huozhe women yao liaojie chang ping de tese(.) genyao liao jie zhege changing xiying zhemeyangde zu queng(.) zhenyang cai hui bi jiao wanzheng xueng xi? haishisheme?

or we may need to understand the product but what is more important is we need to understand what attracts the target customer. Will the argument be more complete in this way? What do you think?

183 S10 Wojiede::(.) di: er:ge: bushihe:

I think the second reason is not appropriate here.

184 T Hao. Bushiyong. S4 ni xiangyixiang S10 genide jienyi. Tarenwei changpin tese bushi zongdien

OK. the second reasoning doesn't fit here, there is no logical relation. ↑S4! Think about this suggestion. S10, I think it is not appropriate here.

4.6 Summary

In this chapter, the results and analysis of the quasi-experiment are presented. The results contain the both quantitative and qualitative components. Data drawn from the experiment was analysed by different methods.

4.6.1 Peer Feedback and Writing

The results provide evidence of the effects of the 'step' training method and the peer assessment practice. Three aspects of the effects of step training were examined: the usage of steps by students in peer feedback; the quality of students' peer feedback; and the quality of students' argumentative writing. The results were analysed with *t*-tests and ANCOVA to understand differences between the groups. It was found that the 'step' training given to group A provided significant overall advantages over that given to the control group. The outcomes agreed with Min's (2005) results to a large extent. Moreover, it was also found that the difference in the usage of four steps was not uniform between the groups. Group A produced some steps more than others. This was also similar to Min's (2005) results.

In order to improve the validity of the quality criteria, the author developed a method which employed criteria (Table 3-5) to consider the relevancy of the writing and comments concerning the issues identified. The method quantifies the quality of writing and peer feedback using a teacher's marking rather than just measuring step usage. The quality assessment provides evidence that the peer assessment training in group A had produced better quality peer feedback than that in group B. The peer assessment also appeared to have had a significant impact on students' quality of writing. The group with the better quality of feedback demonstrated a significant improvement in the quality of its essays, whereas the group with the poorer quality of peer feedback demonstrated no significant improvement.

4.6.2 Students' Attitudes, Perceptions and Opinions

Questionnaires were distributed to measure students' attitudes and perceptions. The results were analysed using independent *t*-tests and ANCOVA. The design of the questions was based on the theory of planned behaviour. The main findings were as follows:

1. It was found that despite the different levels in the quality of peer assessment, both classes had a strong preference for the teacher's assessment rather than formative peer assessment (see section 4.4.7).
2. The two groups showed significant differences in opinions, perceptions and attitudes concerning peer assessment. Some of these differences correlated statistically with differences in peer assessment performance between the groups. Students' perceptions and attitudes towards peer assessment were influenced by their training in peer assessment. Contrary to the author's expectations, the behaviour of the students in the control group whose peer feedback was poor in quality was more influenced by the peer assessment training compared to the students in the experimental group who gave better peer feedback.
3. The analysis of interview data has revealed more detail about some detailed opinions of the students. From the themes raised by students during the interviews a few common issues could be identified such as the importance of language proficiency, learning needs, time distribution during peer assessment and levels of thinking; these issues to some extent reflected the details of empirical experience in peer assessment. Discussions based on these findings are included in the following chapter.

4.6.3 Classroom Interaction

Classroom interaction and discourse were analysed using Bernstein's formulation of socio-cultural learning theory. The concepts of classification and framing formed the basis of the qualitative analysis. The two discourses of modelling and negotiation were analysed.

With modelling discourse, a clear boundary was established between step usage and writing instruction. For negotiation discourse, a clear boundary between writing instruction for content and logic was made while a weak boundary between step training and writing instruction existed. In the modelling discourse, both regulative and

instructional framing was strong. In the negotiation discourse, two more complex combinations of framing were found when the teacher tried to deal with errors in strong classification.

Movement from convergent to divergent assessment was analysed with typical snapshots of the classroom interaction as demonstrated in different excerpts. Teacher's role changes and framing analysis demonstrated that, by using different roles and framing during the interaction, the student could achieve a different level of learning. Student identity change was not directly observed; however, the evidence showed that changes in identity could have occurred during the interaction and that divergent assessment might have contributed to this.

Chapter 5 Discussion

The previous chapter presented the major results of this study. The author focus now on the following points for discussion around peer assessment training and student learning.

5.1 Results for the ‘Step’ Training for Peer Assessment

The results of this study have provided statistical evidence of the effectiveness of ‘step’ training. So, are the quantitative results of the students’ work proof that the step method is effective for peer assessment? In line with the previous literature, the quantitative results have shown that, with the help of ‘step’ training, the experimental group demonstrated a higher quality of peer feedback and writing than the control group.

5.1.1 Positive Evidence from the Experimental Group

The main evidence concerning the positive effects on this group are the following:

- (1) Step usage in peer feedback.
- (2) The quality of peer feedback.
- (3) The quality of students’ written work.

However, the quantitative results cannot fully explain the reason for these positive outcomes. Furthermore, the quantitative results are not fully convincing due to many practical reasons, such as the lack of strict control over the quasi-experimental environment, the absence of a precise measurements of quality, the lack of repeat measures, and other influences such as the Hawthorne effect / observer effect (Henry, 1958).

5.1.2 Weaknesses in the Measurement of “Steps”

5.1.2.1 Possibilities of Biased Measurement Concern

As shown in the results of this research the use of ‘steps’ was just one means of measuring the quality of students’ feedback. However, as this method was only used with a single group of students, this could be perceived as distorting the results.

An attempt was made to address this potential weakness with a comprehensive quality measurement system so as to confirm the effectiveness of the ‘steps’ training.

5.1.2.2 Missing Measurement

Another inherent weakness of the measurement of steps is the difficulty in accurately recording their use. For example, as shown in Table 4-3, the ‘clarify’ step is used very little in both tasks and groups.

This is consistent with the findings of previous research (e.g. Min, 2005; den Berg et al, 2006). It was noticed in den Berg et al (2006) that the students were accustomed to receiving evaluative feedback from teachers but ignored process- oriented feedback that focused on the writer’s plan and intentions. As a result, students had no process-oriented feedback models to follow. Min (2005) noted that students neglected this step because they assumed that they understood their peers after working with them throughout the semester, sharing the same language and cultural background, and so whatever they read that was written by their peers they assumed they understood, even if it might not have been clearly expressed. In this study, the author noticed that the teacher had suggested that students should use this step verbally. When allowed to use verbal communications, as Liu and Sadler (2003) found, students tended not to write this step down in their worksheet. Taken together, these issues would have affected the validity of the findings if steps measurement had been the sole method of assessing the quality of peer feedback. Therefore, the quality of students’ work needed to be additionally assessed by the marking system as detailed in section 3.7.1 and 4.3.2.

Some degree of ambiguity of the measurements cannot be avoided, despite the best efforts made, such as using a control group, and so lessons have to be drawn from theoretical explanations as well as the other experimental measures.

5.1.3 Theoretical Analysis of the Steps

To properly analyse the effectiveness of the learning process, two theories were selected which were appropriate to the research: the theory of planned behaviour and the socio-cultural theory of learning. The theory of planned behaviour is a psychological theory rarely applied to studies of learning behaviour in EFL peer assessment. It emphasizes the impact of human perceptions and intentions on behaviour. In this case, the author wanted to examine the perceptions of students which reflected the results of the training. The perceptions and attitudes were measured and are analysed in Chapter 4, and the implications of these results are further discussed in later sections.

Socio-cultural theory emphasizes that, to achieve effective assessment for learning, teachers and students must have a complex interaction throughout a process of alternating framing and classification (Pryor and Crossouard, 2008). As explained in Chapter 4, the socio-cultural factors have been referred to in explaining the detailed interactions. The implications of these results will be further discussed in the following sections.

5. 2 Students' Perceptions and Attitudes

The theory of planned behaviour (TPB) states that the perceptions and attitudes of a person will form intentions under the influence of various external and internal factors and will ultimately impact on the person's behaviour. Hence, students' perceptions and attitudes should have an impact on peer assessment practice.

5.2.1 *Types of Beliefs*

According to TPB there are varying types of beliefs. The recording of perceptions and attitudes using the questionnaires in this study was not intended to distinguish between these different types of beliefs which would be unrealistic and would not significantly contribute to the results), so the change reflected in the measurements is a combined result.

5.2.2 *Measurement of Psychological Factors*

Changes in opinions, perceptions and attitudes were measured in each group. The pre-Q and post-Q measurements were taken at different points in time, meaning that neither of the questionnaires could accurately capture the precise state of the factor throughout the duration of the training and peer assessment practice. However it would not be possible to take such measurements more frequently during the training because doing so would lead to excessive disturbance to the classroom teaching.

The pre-Q results primarily reflect the students' past experience. The post-Q results, on the other hand, included the perceptions which originated from a range of sources including the training and experience of peer assessment experience. To understand the perceptions originally held by students, we need to look at the pre-Q results and for perceptions held by students after the sessions; we need to examine the post-Q results. To track the changes in perceptions that occurred during the sessions, we need to refer to both the pre-Q and the post-Q results.

Student interviews were also taken into account as a qualitative measure. Although the author considers these as a separate source of information, they were analysed along with the quantitative measures for reference purposes. The interview results were selective because the author only focused on capturing the representative results in order to reflect the overall perceptions.

5.2.3 Students' Perceptions and Peer Assessment Quality

Despite the different outcomes in terms of the quality of peer assessment, both classes expressed a strong preference for the teacher's assessment over formative peer assessment. In sections 4.4.1 and 4.4.7, negative attitudes towards peer assessments are recorded with students confirming in the questionnaire responses that after their experience of the peer assessment practice, they would prefer a teacher's assessment. However, the results in Chapter 4 demonstrate that the 'step' training method effectively improved the quality of peer feedback quality, as well as improving the quality of written work. If the training was a success, why would the students reflect negatively on peer assessment? Moreover, the confidence of students in peer assessment reflected in section 4.4.1 did not increase significantly in either group, with some students also holding complex attitudes towards the teacher's assessment as shown in section 4.4.7. Students in group A became significantly more enthusiastic about the teacher's assessment but students in group B did not.

Overall, the step training enhanced the quality of peer feedback and the quality of peer assessment but also reduced students' preference for peer assessment. This development was considered by the author as non-productive. It should be obvious that if a student has improved his performance in a certain task or is on the right track towards improvement, the student should be motivated or at least have less resistance towards that task. However, it must be kept in mind that by working on compulsory or repetitive training tasks, it is entirely possible to improve performance without increasing the preference for or acceptance of those tasks. But this is not the purpose of formative peer assessment and conflicts somewhat with the ideals of socio-cultural learning, which include to establish the learner's identity so that a higher level of autonomy in learning can be achieved.

5.2.3.1 Discussion of Steps Training Based on Perceptions

As recognised by Min (2006), step training can be inflexible and one of the most important factors in achieving successful formative peer assessment is the movement from convergent to divergent assessment. This requires flexibility and an active discourse of negotiation from both teacher and students throughout the interaction.

However, the steps training introduced a relatively rigid scheme that required the teacher to act in a strictly ‘teacher/assessor’ role. This required the student to accept the steps system and learn to apply them in the context of peer assessment practice with little negotiation with the ‘teacher/assessor’ about the method. As a result of the step training, the students achieved the target with the results confirming that the quality of peer feedback was significantly more advanced than in the group that did not use the steps training. However, this success led to a concern which undermined the negotiation. As previously stated, successful formative peer assessment would be a form of assessment for learning which invites the students to become involved in the assessment criteria. Peer assessment without the active involvement of students would fail to promote the complex shift of power, and the movement in assessment process and identity and role change. The conditions for successful peer assessment were not met in this context. From the students’ perspective, if peer assessment was a process that only involved a fixed series of actions pre-determined by the teacher, then the student would naturally rely on the teacher to provide other guidance for the very reason that the teacher had historically provided such guidance, including the ‘steps’ as a means of producing effective peer feedback. Although the students had some experience of peer assessment, their work was effectively still being evaluated by the criteria set by the teacher or from textbooks.

The techniques used in the ‘steps’ training were well designed to help the students to produce quality feedback. However, this feedback, in a sense, was produced by the ‘steps’ instead of the students. If this experience had dominated the peer assessment practice, then it would be natural that the students were able to provide quality peer feedback, but still retain a strong reliance on the teacher’s assessment. During the post course interviews, the students mentioned that their English proficiency was another barrier for them in engaging with peer assessment, thereby highlighting a further disadvantage of the ‘step’ methods and confirming that the students still had a strong preference for receiving assistance from the teacher, even though the course had

finished. This fact suggests that the learner's new identity was not successfully established.

5.2.3.2 Patterns of Beliefs

The two groups showed significantly different approaches towards peer assessment practice, reflected in their different behavioural patterns. The differences shown between classes could be hidden in the ways they work on peer assessment as a group and not reflected in an individual student's performance or the average performance of the class. Since the theory of planned behaviour helps to establish the link between psychological and behavioural measurements, it is possible to identify the measurement of the psychological perceptions as the driver behind the differences in behavioural patterns. Such measurements can unlock the underlying differences between groups.

As stated in sections 4.4.1 to 4.4.8, the measurement of perceptions revealed significant differences between the two groups. According to the theory of planned behaviour, these significant differences between the perceptions could be predicated to impact on their behaviour, which was found to be the case. It was observed that the two classes displayed differences in the way they performed peer assessment in ways which differences could be linked with the measurement of their perceptions. Hence, this could offer a new line of approach in researching classroom peer assessment.

5.2.3.3 The Influence of Psychological Factors

It was assumed that the difference between the training methods was the most significant difference between the two classes. The variations in training obviously had the most significant impact on the results for the peer assessment outcomes, as shown in section 4.3. It was also assumed that students' opinions had been influenced by their specific training in peer assessment. However, some of the results appear to contradict these assumptions.

As shown in section 4.4.4, both groups strongly agreed with the statement that the training they had received provided sufficient help with peer assessment. However, the training and the peer assessment outcomes in these two classes were different due to the groups not knowing what form of training was provided to the other. They would evaluate the training purely through their own experience. Hence, the perceptions of the

two groups may not be strictly comparable. In this case, the change of the perception within the group through the course needs to be analysed.

As shown in section 4.4.9, contrary to the author's expectations, the behaviour of the students adjudged to have produced peer assessment of the lowest quality was most influenced by the peer assessment training. On the other hand, the students with the better peer assessment quality were least influenced by their perceptions of and attitudes towards peer assessment. This verdict was drawn from the results of Group A which had just one post-perception and two pre-perceptions that was significantly correlated with the peer assessment results. In contrast, Group B had six post questions that were counted. These measurements represented the perceptions most strongly influence by the training as they were made on completion of the training / peer assessment.

A simple count shows that Group A raised only half as many psychological factors as group B. This leads to a reconsideration of the concerns raised in section 5.2.3.1; namely, that the training with the 'steps' performed well but that the quality of peer feedback and writing was less dependent on the students' own perceptions. In group B, however, the peer assessment was accompanied by confusion and lack of direction without the help of the 'steps'. Hence, the results may have been more dependent on the students' own perceptions.

In section 4.4.9, it was shown that the perceptions which impacted on the final results showed different patterns. None of the significant perceptions in the two groups overlapped with each other, but these different perceptions contributed to the different patterns in the peer assessment results. This suggests that the measurement of perceptions measurement could capture the different patterns of perceptions which could contribute to peer assessment performance. Due to limitations of space, however, further analysis of the details of student perceptions and their possible complex impacts are not considered in this thesis. However, this could be reported in further publication.

5.3 Socio-Cultural Learning and Peer Assessment

The major purpose of this thesis was to bring together the training of peer assessment to an ideal practice as described by the socio-cultural theory. In the previous sections, the deficiencies of the training methods have been pointed out without considering the underlying mechanism of 'effective' peer assessment training. This section discusses the ideal practice of 'effective' peer assessment training. Through this discussion, the

results to report above can indicate the correct direction for ‘effective’ peer assessment training to take.

Before conducting the present research, there appeared to be a conflict between the author’s empirical experience and Bernstein’s (1996) formulations. The author initially agreed more with Min (2005) that effective peer assessment could be better achieved through steps training. A procedure used for argumentative writing was thought likely to be the best direction for effective peer assessment feedback, since the two methods share many essential factors such as critical thinking and logical deduction.

However, the author’s own practice in this area in the classroom was never enjoyable. Students raised issues which mainly concerned their level of proficiency in English and they usually failed to see the benefits of the training on their thinking and logical skills. Some of the students regarded the training as inappropriate, making comments similar to these expressed in interviews in this research. The following may be underlying reasons for such negative feedback.

5.3.1 Difficulties in Student Identity Construction during Step Training

The ideal model of formative peer assessment as described by Pryor and Crossouard (2008) provided some direction towards improvement. However, as shown in section 4.6.3, the ideal was not easy to achieve, with or without a well-designed procedure of peer feedback training.

5.3.1.1 Problems with the Teacher’s Workload

In most cases during the experiment, although movement between convergent and divergent assessment could be detected, the students were still unable to provide either an echo style of response, a convergent stamp answer or a divergent miss-by-miles answer. The teacher kept control of the class by using strong regulative framing and authority in the class was never handed to the students, even during ‘negotiation’ which is regarded by socio-cultural theorists as a very important process in identity construction. The teacher frequently failed to provide sufficient instructions to the students during the interaction, which left them confused.

The difficulty of high workload faced by the teacher seemed to be due to conflicts among three areas of interest.

1. Writing instruction. As introduced in Chapter 2, the course had a teaching target to meet.
2. Steps training. The teacher was also preoccupied by having to provide instructions about the steps.
3. Peer assessment criteria negotiation.

In the peer assessment practice, the teacher had to manage these three requirements simultaneously and try to administer them all. If the writing instruction was removed from the equation, the situation would improve, but not significantly. The steps training and negotiation would still mean that the teacher was multitasking. This may explain the occasional hesitations of the teacher during the training when struggling to decide on the appropriate action. As a result, the essential learning construction interactions all appeared in the excerpts, but not consistently as envisaged.

As shown in section 4.4, the interview results generally revealed negative opinions towards the peer assessment. The students found the task hard to understand and there were some negative opinions expressed about the training, such as that it was not well handled and clearly targeted.

This suggests that the workload involved in peer assessment was beyond the capabilities of the teacher, although it could be expected that development of skills in the teacher's peer assessment sessions might improve the quality of future practice.

To address these concerns, it is proposed here that classroom practice using convergent and divergent assessment should be divided into different sessions. By doing this, the teacher could deliver one task at a time instead of multitasking.

Another potential improvement could be to divide the class into smaller focus groups, creating more space for students to engage with their peers. This would also benefit the teacher, who could deliver the modelling discourse to smaller groups.

5.3.1.2 Improving Student Confidence Levels

Another issue reflected in the literature (Min, 2010; Hu 2005) as well as in the results above is the student concerns about their English proficiency. It is a simple fact that students are not confident about becoming peer reviewers when there is generally low

confidence. So, the confidence level is a barrier to constructing the appropriate student identity for peer assessment.

Min's (2010) and Hu's (2005) students had significantly superior skills in English than the students in this study (see Chapter 2), but the advantage in English proficiency cannot eliminate the concern about confidence levels. Hence, if teachers want to address this problem in the training, they may face two choices: one is to improve student English proficiency levels, and the other is to improve the students' confidence by other methods, such as encouraging them by using special preparation sessions. For example, when introducing the peer assessment criteria and steps, the present author included an interactive activity explaining to the students likely problems with the expected peer assessment, and examples demonstrating how to give good peer feedback. By doing this, the students could become more familiar with the peer assessment activity, and their confidence levels may be increased. Moreover, it is further suggested that the teacher could also demonstrate to the students that, in terms of feedback on content, they might be able to give comments that are as useful as, if not more useful than, those provided by the teacher. Furthermore, it could be emphasized that and b) that feedback on content is as useful as, or more useful than, feedback on language. By doing this, the students' confidence levels could be further improved which will also help to establish the learner's identity.

5.3.1.3 Micro and Macro Aspects of Writing in Peer Assessment Training

Finally, both Min (2005) and Hu (2005) criticized previous claims that peer assessment should focus more on the macro aspects of writing and pay less attention to the micro aspects such as in Nelson and Murphy, 1992; Paulus, 1999; McGroarty and Zhu, 1997). They claimed that the micro aspects of writing should be treated as equally important because this is one of the essential tasks of their teaching. The author agrees with this claim in the context of EFL learning. However, this issue can be considered from the following point of view.

Min and Hu focused on the micro aspects of writing because they were driven by the students' needs. It is almost impossible for a responsible teacher to ignore the student's essential needs. Improving the micro aspects of language is one of the essential needs of EFL students. For example, if during peer assessment training a student raises a question about grammar, it would be difficult for the teacher to ignore this question.

But, answering this question may trigger following questions. The teacher will face a choice between remaining focused on the macro aspects or answering the questions. Min and Hu's suggestion is to focus on both aspects instead of just one.

For a student in a class with relatively high English proficiency, the frequency of issues revised about micro aspects of language may be within an acceptable range. However, for the present author's students who have much lower English proficiency, the focus was drawn principally towards the micro aspects. Hence, the peer assessment training method has to be modified based on the actual situation. For example, the author modified the training to include an introduction to all of the micro and macro aspects of writing and what to expect in peer assessment. It was expected that the clarification of the micro and macro aspects of writing can help the students to understand the peer assessment task, so that they would be able to anticipate the issues encountered in the peer assessment. The outcome of this training will be an interesting topic for future research.

5.3.1.4 The Value of Exploratory Talk in Peer Assessment Training

One important aspect not discussed in the previous literature is the type of talk used in formative peer assessment. It is important to use the most appropriate mode of talking during different stage of the formative peer assessment. The teacher's talk will be the medium for meaningful and transformational learning. Failure to attend this issue may result in many interventions making little progress (Mercer and Hodgkinson, 2008). As a characteristic of formative peer assessment, exploratory talk is essential to the divergent build up of the assessment (as mentioned in section 5.3.2). However, the value of exploratory talk is more than a symbol of divergent assessment; instead, it is important from a methodological perspective.

The step training is introduced to students to help them understand and produce peer feedback. The training is quantitatively proven to have achieved improvement in students' work but it also created problems such as those which have been recognized in sections 5.2.3.1 and 5.2.3.3. This result has identified the function of intensive and tangible strategies that have worked efficiently but only at the surface level. The value of exploratory talk which takes place and works as a feedback rich process in peer assessment is not recognized in this type of training or discussed in previous research works. In such students works (i.e. Min 2005, Hu 2005), the lack of a socio-cultural

perspective and the major focus on quantitative evidence may mislead practitioners and teachers to drawing the conclusion that simple and rigid training methods might represent a universally effective procedure. In reality, when applying these procedures in teaching practice, the superficial peer assessment will result in negative experiences for both teachers and students. Exploratory talk is then extremely valuable to avoid this situation, for the following reasons:

1. It allows mutual feedback between teacher and students, which will avoid misunderstandings by the teacher about the students' actual development. For example, a rubber stamp feedback could easily be identified if the teacher uses the talk to identify the actual meaning of the peer feedback.
2. Exploratory talk will make the teacher more aware of the students' actual needs during the training. It will also help to identify any gaps between the teacher's training plan and the students' demands. Especially in formative peer assessment, where the student's need could be diversified. The exploratory talk will help the teacher to realize the difference between formative and summative assessment and remind the teacher to adjust their roles.
3. The students will also benefit from the exploratory talk. Moreover, they will be aware of the different power relationship in this talk. It will help them to understand the core value of the formative peer assessment and to adapt their role.

In conclusion, exploratory talk is an essential characteristic of formative peer assessment and the importance of the talking environment for successful peer assessment training should be made explicit to teachers.

5.3.1.5. Discussion Based on Observation of Student Identity during Step Training

As introduced in chapter 2, the best achievement of peer assessment training is to develop the student's identity, which reflects the fact that they have fully developed their peer assessment skills and established the identity of a peer assessor. Hence the students could work on peer assessment on a highly autonomous basis. However, before the ultimate target of the peer assessment training could be achieved, the status of the student's actual identity could not only determine their performance during the peer assessment tasks, but may also affect their opinions about and perceptions of the practice.

Pryor and Crossouard's (2010) summary of the different levels of learning during formative assessment included: completing the task in hand, making sense of meaning, thinking of improvements and invoking the learner's identity. If a student's identity is defined by whether the learning has achieved a full identity change, there will only be two identities, one as student, and the other one as peer assessor. However, based on the observations made in this study, it is predicted that only a fraction of students can achieve the level of becoming a fully mature peer assessor as a result of a prolonged period of training and practice. Hence it will be useful to detail the identity status of the students whose new identity is not fully invoked. Pryor and Crossouard(2010) propose four different levels as: identity 1 where the student is only able to perform the task at hand, through to identity 4 which represents the fully invoked learner, and the following discussion matches the feedback from students in the present study with these levels of identity.

For identity 1, which corresponds to the learning stage of completing a task in hand, the student may be able to produce work based on the explicit tasks, but shows limited ability for a task outside the scope of direct instruction, such as negotiation. For students with this identity, one of the obvious problems is the passivity of their actions. Such passivity may not be a major problem for summative assessment; however, it will become a major threat for the formative peer assessment. Students with this identity will find themselves experiencing difficulty in collaborating with other students. Firstly, they may not be able to correctly understand the meaning of others' work; and hence they may not be able to identify problems in the work. Secondly, students with this identity may also lack the skills to communicate effectively with their peers. Naturally, students with this identity will have difficulty in producing genuine peer feedback. Moreover, students with this identity may consider completing the peer feedback task as an obligation during the training session. Hence, they may give 'rubber stamp' feedback or feedback copied from training examples is commonly expected. The major issue for students with identity 1 is that they have neither the appropriate ability nor the correct identity to carry out the peer assessment task.

For identities 2 and 3, the student may be able to produce acceptable peer feedback and provide reasons to persuade their peers; however, their identity is still not that of a true peer assessor. Their problem is that, although they have demonstrated their capability in peer assessment during the training, this ability will be limited to the training environment. They may not be able to perform peer assessment as independent

assessors in a different situation. The problem these students may have is that they can develop the appropriate manner but not the identity of become a peer assessor.

Students with identities 1 to 3 will all have a common tendency of relying on the teacher to set up the peer assessment tasks and topics, and this is reflected in the questionnaire results. Subsequently, they will reject the idea that the teacher will not be the leader of the practice, and this is again reflected in the questionnaire results. Moreover, students with identities 1 to 3 will rely on instructions and guidance from the teacher at different levels. The general tendency is that clear and rigid instructions will be preferred rather than divergent criteria, because these will be easier to follow. Since most of the students in the present study did not have identity 4, it is not surprising to find that questionnaire results and interview feedback are dominated by the conflicting outcomes.

The discussion above considers the influence of identity status on peer assessment. However, another tendency was also noticed which is worth mentioning, which is the influence of conflict during peer assessment. During this study, obvious and direct conflicts between student peers were rare. The reason for this may also lie in the clash of students' identities when resolving conflicts. Despite their development during the training, the students were generally not comfortable taking on the assessor's role. One of the reasons is related to agency and identity transition, as discussed above. However, the author noticed that the students' discomfort was partially caused by the fear of causing conflict between peers, since no training was provided to cover the issue of potential conflict and the skills needed to resolve. In this case, it was difficult for the students to achieve complete identity development. As mentioned in chapter 2, previous research by Connor and Asenavage (1994), Nelson and Carson, (1998) and Kamimura (2006) looked at this issue in terms of a cultural background and came up with contrasting results. However, if it is assumed that the fear of conflict is a common human reaction, then cultural background can be dismissed as an influence in this study. This therefore suggested that if the risk of conflict could be clarified as part of the peer assessment training and that conflict avoidance or resolution skills were included in the training, then the fear of conflict could be reduced. The issue of conflict would then be reduced to a problem of agency which prevents students' identity change.

For example, it is apparent in most of the excerpts the teacher's assessor role is to evaluate the right or wrong answers, to talk about the answers, and to provide

suggestions. Therefore, this role suggests the need for qualities of criticality and authority which students are not very comfortable with or capable of taking on themselves. During the interview, when asked for an opinion about criticality between peers, S4 commented that everyone has different criticality and it is hard to say who is right or wrong. He pointed out that the role of being a peer assessor can be dangerous if both exercise their criticality and authority consciously or unconsciously. This strong hierarchical relationship suggests conflicts between peers and jeopardizes their friendship (see, for example, excerpt 3). To students, the new identity means that they run the risk of embarrassing themselves or their peers. As a result, even if the students have the skills necessary for peer assessment, the majority of them choose to remain silent in the discussion. In facing peers who exercise their powers and decided to take action, students may compromise easily to keep the peace. For example, in excerpt 6, S10 exercised his authority by questioning the other's answer, in an exact imitation of the teacher's action, but then S4 turned away from the conflict and the questions by saying 'OK, OK' quickly to end the conversation. For these students, their first reaction to a negotiating question is to avoid the use of power in the peer assessor role and to avoid conflict. This may help to explain why the students' perception of their ability in and attitudes towards peer assessment remain similar before and after the intervention. That is, even if the students had developed peer assessment skills, they may not have felt able to adapt to their peer assessor role due to the lack of, for example, conflict resolution skills. As a result, the students may either choose not to change or become more conservative about their new role. Hence, identity change was not accomplished.

Therefore it can be proposed that conflict awareness sessions which include conflict resolution skills should be included in the peer assessment training to prepare students to take on the peer assessor identity. Secondly, teachers should understand that the agency required by the student to facilitate the identity change may exceed the scope as planned. For example, the skills related to conflict were not originally considered for the training in this work, but they may be very important in allowing the student's identity to change. This also underlines the importance of flexibility and adaptability in teachers, who have to be able to understand their student's needs and adjust the discourse accordingly.

5.3.2 Teacher's Role

According to Pryor and Crossouard (2008), formative peer assessment ideally requires

teachers to be flexible and to adapt their roles according to the needs of the students' learning in the classroom. The focus of the teacher's role should, ideally, suit the movement from convergent to divergent formative peer assessment. This presents new challenges. Previously, the teacher was expected to act like a chairman. But now the teacher has to run like a football referee, coach like a manager or even join the team to play when required. This is difficult for most teachers who have previously been used to the traditional teacher's role. However, it need not be so difficult if the teacher adopts the correct approach.

The approach used by the teacher should be to be well prepared, equally treated and clearly directed:

1. The teacher has to be well prepared for divergent assessment. This is probably the most problematic aspect of formative peer assessment, but without it, learning will not be effective. This requires the teacher to have a good grasp of general knowledge, as well as being able to anticipate students' questions. If not, the teacher would tend to use their authority to re-direct the topic back towards the teacher's preferences (as shown, for example, in excerpt 3). This could undermine the outcome of divergent assessment.
2. The teacher has to be able to treat the students as equals even though all may have recognized that the teacher has more knowledge on the subject. In the measurements of perception (see section 4.4.7), the students showed an increased preference for teacher assessment even though the quality of their own peer assessment improved. This suggests that the students were not prepared for the peer assessment and did not realize that, they played an equally important role in the learning process. Hence, it is important for the teacher to prepare the students with a renegotiation of teacher and learner relations (Hu, 2005; Prior and Crossouard, 2008).
3. In general, the teacher has to keep the class under control. When an individual student requires guidance, the teacher would have the choice to deal with it immediately or not, depending on the best interests of the class. Formative peer assessment seems to offer a great amount of flexibility, but if the teacher does not provide clear direction by using their authority appropriately, the students may become confused (as shown in excerpt 3) and this could lead to distress, further reducing the student's enthusiasm for peer assessment.

4. According to Pryor and Crossouard (2008), the role of the teacher in the process of formative assessment has to be flexible in order to allow space for negotiation to take place. In divergent assessment, negotiation would be especially likely to occur when the teacher takes on the role of knowledge expert and learner. This breaks the IRF pattern to allow real and meaningful discussion about the criteria used, and also allow relations of equality to develop between participants which enable the flow of the discussion. In this training course, however, the teacher mainly took the role of teacher and assessor rather than knowledge expert and learner. This resulted in a hierarchical relationship among the participants during most of the classroom interaction. This imbalance led to the increased appearance of IRF patterns and reduced the chance of opening up space for real and meaningful negotiation about the criteria. In consequence, students were trained properly to follow the steps and guessed what the teacher desired of them, but did not develop an understanding of the criteria for giving peer feedback, which is considered to be the most important characteristic of an independent peer reviewer.

The observation of the above interaction has provided a possible explanation for the apparent conflict in the results between the qualities of the students' work and their perceptions of peer assessment as discussed in 5.2.3. Torrance and Pryor (2001) have identified similar issues in formative peer assessment classrooms which are considered to pose major difficulties for the practice. In the light of the experience gained in this study, the findings support the conclusion of Torrance and Pryor that the lack of role change in the teacher to that of a knowledge expert and learner is likely to be due to a lack of experience in divergent assessment practice. It would be useful for the teachers to recognize the importance of their roles in divergent assessment as well as to recognize the characteristics of the talk environment needed for divergent assessment. One kind of such talk is exploratory talk, which allows the learners to "try out ideas, to hear how they sound, to see what others make of them, to arrange information and ideas into different patterns" (Barnes, 2008). This is obviously different from the ordinary 'presentational talk' used by teachers in their traditional role. Hence, the use of exploratory talk could be a critical skill to allow the teacher's role to change in divergent assessment. It is also important for the teacher to

realize the importance of using exploratory talk in the practice of formative peer assessment.

In this research, a typical example of exploratory talk appears in Excerpts 5 and 6, where the teacher engaged in an in-depth discussion about the strategy for making an argument. The teacher explored the opinions of the students and allowed prolonged discussion between them while maintaining equal power relations between the students and the teacher during the interaction as well as among the students during peer assessment. This resulted in a highly divergent conversation during which the students managed to produce one of the most extensive sets of oral comments found in this research. During the conversation, the teacher's role was not obviously that of a teacher; instead, there are occasions where different students interrupted the discussion with clarifications, reflecting their strong motivation to express and test their thoughts. The level of complexity of the students' comments here is also encouraging. The teacher did not act as the leader in presenting complex ideas and reasoning. The students were trying to explain or clarify the complex reasoning behind their conclusions. Based on this observation, it is suggested that exploratory talk can help divergent assessment and the generation of comments. It allows a change in the teacher's identity and in the meantime helps to establish the learner's identity.

Chapter 6 Conclusion

In this thesis, training in formative peer assessment was examined in an average level EFL writing class. The author of the present study identified gaps in knowledge between the relevant literature and theoretical ideals. A quasi-experimental and case study approach was adopted to examine a peer assessment training method proposed by Min (2005). To conclude this thesis, the research questions posed at the beginning of this thesis are first revisited:

1. What was the nature of peer feedback provided in both groups?

In the quasi-experiment peer feedback was collected from two groups of students: the experimental group (Group A) and the control group (Group B). Significant differences were found in the quality of their feedback.

- a. To what extent did the peer feedback follow Min's four steps?

When measured by the use of steps as proposed by Min (2005), the students in group A had completed significantly more steps in their feedback than those in group B. The average number of steps contained in each piece of feedback in group A was significantly higher than in group B (as shown in section 4.3.2). Moreover, an assessment of the quality of the peer feedback also demonstrated that the group following Min's steps produced higher quality feedback (as shown in section 4.3.3).

- b. What was the quality of the students' argumentative writing?

The quality of students' writing after the peer assessment was significantly influenced by the quality of students' peer feedback. The experimental group improved the quality of their writing after the peer assessment compared with the control group (as shown in section 4.3.4).

- c. What were the students' attitudes towards and perceptions of peer assessment before and after the training?

The research measured students' attitudes and perceptions towards peer assessment before and after the training by using questionnaires. The attitudes towards and perceptions of peer assessment before and after the course were compared and the differences investigated using t-test analysis. It was found that the patterns of students' attitudes and perceptions in the two groups showed

clear differences, and these were then analysed in terms of the quality of peer feedback and writing using ANCOVA. This analysis suggested several statistically significant associations between the quality of peer feedback and writing produced and the attitudes and perceptions of the students.

Overall, a complex result had been demonstrated. Although the step training significantly improved the outcomes of the peer assessment, concerns were raised from the measurements of attitudes and perceptions that the rigidity of the training method could undermine the students' confidence in and motivation towards peer assessment.

2. How did the teacher model and negotiate to shape students' peer feedback?

It was assumed that effective peer feedback training achieved through the use of two types of discourse: modelling discourse and negotiation discourse. Throughout the use of these discourses, the teacher and the student went through a complex series of interactions involving alternations of framing and classification, the teacher's role play and the students' identity construction.

a. Based on Bernstein (1996), were the discourses of modelling and negotiation explicit in class?

During the step training, the modelling discourse was explicit. However, negotiation discourse was not straightforward. The teacher could not relinquish the authority to encourage equal negotiation with the students. The students were not adequately involved in the negotiation process and as a result did not complete their learning. When the students did not have a complete understanding of peer assessment, they were still able to follow the step training well during the basic peer assessment task. However, they could have been less successful with a more difficult peer assessment task (as shown in section 4.3.3).

b. How did the teacher engage students in the discourses of modelling and negotiation?

The teacher should engage students using the technique of suitable framing and classification setting and flexible role play. The importance of these two essential strategies was demonstrated in section 4.5. Moreover, the teacher should provide sufficient modelling discourse by using strong framing and

classification, during which the teacher should adopt the strongest role to use authority to deliver the modelling. This will improve the effectiveness of this discourse by clarifying the scope of the training and the tasks involved in it. It has also been suggested that a session with interactive activities demonstrating how to provide peer feedback and what to expect in peer assessment could help to build up students' confidence in peer assessment.

During the negotiation discourse, the teacher should use more divergent approaches and should be able to help the students construct their identity while going through the negotiation process as equals. As pointed out in the socio-cultural theory of formative peer assessment (Pryor and Crossouard, 2008), the negotiation process is critical in establishing the student's identity as a peer assessor. The observations of classroom interaction in this research confirmed this by presenting evidence of the interaction between teacher and students' during the training. It was shown in the analysis that negotiation is critical for the student to establish their roles and to take stances during the interaction, so that the student could have the opportunity to establish their identity in peer assessment. It is also important to maintain the motivation of students as well as their confidence in the process of peer assessment practice. During the discussion, it was suggested that enhancing the student's motivation and confidence levels should be one of the major functions of the peer assessment training. The reason for this is that self-confidence in peer assessment is a crucial factor in constructing the peer assessor's identity. It is suggested that the peer assessment training should be designed especially to improve the student's motivation and confidence level.

Having answered the research questions of this study, the following general conclusions can be drawn:

1. Training methods such as the 'step' training proposed by Min can be effective in certain contexts. Students can improve the quality of their peer feedback through attending such training programmes. The underlying reason for the effectiveness of such training methods can be attributed to the appropriate application of instructive framing during the modelling discourse, which gives the students a clear idea of what to expect and what to do during the peer assessment.

2. The drawback of such a training programme was the scarce use of negotiation discourse. During the administration of the training programme, the students had insufficient opportunity for negotiation. This deficiency can undermine the student's learning cycle as well as their motivation towards peer assessment.
3. The rigidity of such a training programme can have a negative impact on formative peer assessment. Problems such as the teacher's multitasking could affect the success of the modelling discourse itself and improvements should be made to promote the flexibility of the teacher's roles as well as the framing conditions.
4. Student's perceptions contained sophisticated information regarding the peer assessment practice. Statistical analysis helped in identifying the most significant perceptions concerning the peer assessment. The analysis revealed different perception patterns which can be linked to the peer assessment outcomes using the theory of planned behaviour. This suggests an alternative approach to researching peer assessment practice.

To achieve effective peer assessment, a training method with clear modelling discourse would be necessary. However, it is concluded that the true effectiveness of peer assessment could only be achieved through a rather complex process of interaction between the teacher and students and among peers. Due to the complexity of the peer assessment training, an ideal simple and rigid training method for peer assessment will not be universally effective. The training method has to be modified to meet the actual needs of the students. Socio-cultural factors should be carefully considered during the planning of peer assessment training in order to improve the effectiveness of classroom interactions and student's identity construction, so that the effectiveness of the peer assessment training could be improved.

Appendix A Training steps suggested by Min (2005)

Training steps suggested by Min(2005)			
Step	Definition	Examples of comments	Instruction and examples on the worksheet in this study
Clarifying the writer's intention	Students-peers try to get further explanation of what writers have said or what is not clear to them in the essays (e.g., an unknown term, an idea)	What do you mean by that? Why will she agree to the rule because [of] her father's death? Why will she agree to the rule because she wants everyone to be healthy?	Please write down the sentence or ideas need work. Ask your peer questions to clarify the part you don't understand or you think is inappropriate. For example, "She also concern economic pressure problem of her family. Especially, she plans to have a baby, the coast must be a lot in the future.' Why did Jessica need to think about this problem?
Identifying the problem	Students-peers cite a problematic word, phrase, sentence or cohesive gap	I think you need to point out the relations between the example and the policy.	Please write down the reason why you think there is a need for revision.
Explaining the nature of the problem	Students-peers explain why they think a given term, idea, or organization is unclear or problematic, which should or should not be used in the essay	If you did not explain the relations between them, it is difficult for the readers to follow your logic and the logic is weak too.	Please state why you think the problem should or should not be used in the essay.
Making specific suggestions	Students-peers suggest ways to change the words, content and organization of essays.	If you are trying to say that Jessica will agree to the rule because [of] her concerns of [about] health maybe you can say it in this way: Jessica may agree with the new rules from [the] company. The reason why she agrees is because she wants everybody [to be] healthy, and both smoking and second-hand smoking are not healthy. One typical example will be her father who died from lung cancer because of smoking.	Please provide your specific suggestion for the changes. (If you were the author how would you change it?) For example, if you are trying to say that Jessica will agree to the rule because of her concerns about health maybe you can say it in this way: Jessica may agree to the new rules from the company. The reason why she agrees is because she wants everybody to be healthy, and both smoking and second-hand smoking are not healthy.

Appendix B Peer assessment worksheet

Group A : Bill or Kwame?

Group _____ Your ID _____ Name _____ Date _____

Part A General review:

1. 你最喜歡這篇文章的哪個地方?

2. 請看看同學的文章裡有沒有這些東西，如果有就打勾。(請確認同學寫的大方向是對的:文章重點應該是寫出他/她不同意選擇 **Bill** 或 **Kwame** 和為什麼他/她不同意這個選擇的三個原因。)

文章結構	✓	範例
開頭 ✓✓		寫出開頭:I just learned about that you decided to appoint Bill as your apprentice. (✓) 寫出他/她不同意這個選擇:I know you feel strongly about Bill, but (✓)
理由 1 ✓✓		第一個理由:First, Bill is just too anxious about his work. (✓) 解釋第一個理由: While he is anxious, he tends to accomplish everything himself rather trusting his team members to do it. No team work can be successful if the leader is not giving the responsibility away. (✓)
理由 2 ✓✓		第二個理由:Secondly, his personality is too hasty for a big company. (✓) 解釋第二個理由: Since Mr. Trump has a large company; he needs someone who is calm and stable. Bill is described by his team member as too hasty and anxious. His characteristic of lacking patience might put the company in trouble. (✓)
理由 3 ✓✓		第三個理由:Thirdly, by looking at Bill's performance, he is a winner but not a good team player. (✓) 解釋第三個理由: He always did well on his part but never gained trust or friendship from his team members. This indicates that he might have problems when he has to deal with the relationships among colleagues. His colleagues may not follow his lead or help him at work because of his winning attitude.
結尾✓		重申不同意的立場(或給予建議):Therefore, by taking all these into consideration, it is wise to reconsider the candidate. (✓)

3. 請仔細閱讀同學的文章，並在下面勾選這篇文章需要改進的地方。

_____	開頭沒有主旨句或這個主旨句沒有明確說出為什麼要寫	_____	拼字文法標點符號的錯誤還需要多次的修改
-------	--------------------------	-------	---------------------

	這篇文章		
——	結尾沒有提出建議	——	文章的邏輯沒有表達清楚有意義
——	文章結構不清楚，不便於閱讀	——	舉的例子不貼切文意
——	文句一直重複用詞	——	文章無法吸引讀者閱讀
——	長度沒有達到 80-100 字內的規定	——	其他:_____

Part B Critical Review:

1. 作者的寫作目的是什麼?

2. 你同意作者的論點嗎? 為什麼同意或為什麼不同意?請解釋。

3. 你覺得這篇文章論述的清楚有力嗎?
1.....2.....3.....4.....5
4. 請寫出要改的地方,並把要改的句子或論點抄寫下來。可用問題的方式，請對方釐清你不懂或覺得不合宜的部份。

5. 請寫出問題點 (為什麼你覺得這裡需要修改),並解釋這個問題點的重要性(如若論點不夠強而有力或解釋不清楚，可能影響文章說服力。)

6. 我的具體建議(如果是你，你會怎麼改?)

Part C Responding to the suggestions:

- 1.請閱讀同學給你的建議，並抄寫下你覺得有用(或是沒用)的建議。

- 2.為什麼你覺得這個建議有用(或是沒用)呢?

The writer's second draft for the disagreement to the candidate choice (第二稿).

Group B: Bill or Kwame?

Your ID _____ Name _____ Date _____

Part A General review

1. 你最喜歡這篇文章的哪個地方？

2. 請看同學的文章裡有沒有這些東西，如果有就打勾。

文章結構	✓	範例
開頭 ✓		寫出開頭: I just learned about that you decided to appoint Bill as your apprentice. (✓) 寫出他/她不同意這個選擇: I know you feel strongly about Bill, but (✓)
理由 1 ✓✓		第一個理由: First, Bill is just too anxious about his work. (✓) 解釋第一個理由: While he is anxious, he tends to accomplish everything himself rather trusting his team members to do it. No team work can be successful if the leader is not giving the responsibility away. (✓)
理由 2 ✓✓		第二個理由: Secondly, his personality is too hasty for a big company. (✓) 解釋第二個理由: Since Mr. Trump has a large company, he needs someone who is calm and stable. Bill is described by his team member as too hasty and anxious. His characteristic of lacking patience might put the company in trouble. (✓)
理由 3 ✓✓		第三個理由: Thirdly, by looking at Bill's performance, he is a winner but not a good team player. (✓) 解釋第三個理由: He always did well on his part but never gained trust or friendship from his team members. This indicates that he might have problems when he has to deal with the relationships among colleagues. His colleagues may not follow his lead or help him at work because of his winning attitude. (✓)
結尾 ✓		重申不同意的立場(或給予建議): Therefore, by taking all these into consideration, it is wise to reconsider the candidate. (✓)

4. 請仔細閱讀同學的文章，並在下面勾選這篇文章需要改進的地方。

_____	開頭沒有主旨句或這個主旨句沒有明確說出為什麼要寫這篇文章	_____	拼字文法標點符號的錯誤還需要多次的修改
_____	結尾沒有提出建議	_____	文章的邏輯沒有表達清楚有意義

——	文章結構不清楚，不便於閱讀	——	舉的例子不貼切文意
——	文句一直重複用詞	——	文章無法吸引讀者閱讀
——	長度沒有達到 80-100 字內的規定	——	其 他:_____
			-

Part B 請寫出任何需要修改的地方，並解釋為什麼要修改。最後給予具體建議。

Part C 請在空白處重寫你的文章。

Appendix C Letter of consent for the participants in this study

Dear students,

This research intends to discover the relationship between peer reviewing and your writing production. In the period of data collection, the researcher will video and audio-record the class, distribute questionnaires and conduct interviews. All the data collected in this period of time will only be used by the researcher and your teacher. Your information will be kept confidential and destroyed after five years. If you agree to participate in this research and authorize the use of the collected materials, please sign your name below.

各位同學,

本研究主要了解同學互評與寫作的關係。在本研究裡，研究者將會實施課堂觀察、課堂錄影、錄音、問卷調查和個人的面談。採集到的所有資料，您的意見與個人資料也僅限於本研究與授課老師使用。您的個人資料也將受到嚴格保密，並於五年後銷毀。如果你同意參予此次研究並授權以上資料的使用權，請在下方處簽名。

Appendix D Student Interview Questions

A. After session interview

1. What do you feel about the class?
2. What do you feel about peer review? Do you like it? Why? Why not?
3. What's peer review to you? What's the purpose?
4. What kind of feedback do you usually get? What kind of feedback do you usually give?
5. Did you feel uncomfortable/stressed when doing peer review? Why? What did you do to overcome these feelings?
6. Do you think peer review helps you? In what way?
7. What is the difficulty you have when doing peer review? What did you do to solve the problems?
8. Do you feel stressed in peer review or in the course?
9. What do you want to get from peer review? Did you get it?
10. Do you think your peers can do corrections for grammar, content, organization?
11. Do you think the training helps you to produce comments in peer review? Why or why not?
12. Do you think the training helps you to evaluate your peer's comments?
13. Do you think the examples provided to give comments are useful?
14. Do you ask clarifying questions? If yes, in what situation, what did you do? How often in a peer review? (give a number if you can)
15. What do you think is the difference between the teacher's comments and peer's comment to you?
16. Did you change your mind about peer review after you had this training course?
What did you change?
17. Do you think peer review is necessary in an English writing class? Why or why not?
18. Anything you would like to add?

B. After training interview

1. Do you think the teacher should be the reader and evaluator of your writing?
2. Do you think your teacher is very good at motivating you to improve your writing?
3. Do you think your classmates can motivate you to improve your writing?
4. Do you think teacher assessment can be replaced by peer feedback activity?
5. Do you think peer review informs the result of your study?
6. Do you think you cooperate well with your peers in peer review?
7. Do you think you have the ability to peer review the work of others?
8. Do you think the training received helps you to think and give comments?
9. What did your peer feedback focus on? Grammar, content or organization?
10. Are you frustrated when doing peer review?
11. Do you feel offended when doing peer review?
12. Do you think the time allocated for peer feedback activity is enough for you? Why or why not?
13. Do you like peer feedback activity?
14. Do you think you are more capable of producing English writing?

Appendix E Questionnaire Questions

Qs	Statement	Strongly agree	agree	partially agree	disagree	Strongly disagree
Students' perceptions and attitudes concerning teacher/peer assessment						
1.	Teachers should be the readers and evaluators of my writing.					
2.	My teacher is good at motivating me.					
3.	My classmates can motivate me to improve my writing.					
4.	Teacher assessment can be replaced by peer feedback activity.					
5.	Peer review informs result of my study.					
Students' perception of ability to produce peer feedback						
6.	The training I received helps me to think and give comments.					
7.	I think I have the ability to peer review the work of others.					
Students' perceptions of the aspects of writing						
8.	My peer feedback focuses on grammar correction.					
9.	My peer feedback focuses on content.					
10.	My peer feedback focuses on organization.					
Students' perception of peer collaboration in peer feedback activity						
11.	I cooperate well with my peers in peer review activity.					
12.	I am frustrated when doing peer review.					
13.	I feel offended when doing peer review.					
Students' perceptions and attitudes about step training						
14.	The time allocated for peer feedback activity is not enough.					
15.	I like peer feedback activity.					
16.	I consider myself to be capable of producing English writing.					

Appendix F Coding for students written feedback

Mark	Relevance and specificity of formative peer feedback	Students' examples
5	Problems identified correctly complete with reason, complete with valid suggestion	You should point out which company we should know more about and which company will benefit from the investment. If you can point that out clearly in the writing, it will be easier for readers. For example, we need to understand more about Mesa Foods such as their products, the strength and weakness, so it will be better to our company, Omni Inc.
4	Problems identified complete with reason, complete with weak improvement	I think there is no relations between the age of target group and the decision of buying Mesa Food since 14-25 is the target customer of Omi Incl. You can say instead that snack food are popular among children or young group which is the target group of Omni Incl, but we don't understand their way of spending money neither the allowance for the cost. In other words, you should explain first what is the target customer of Mesa Food and their ability to spend and consume on such product.
3	Problems identified complete with reasons complete with invalid suggestion	I think the arrange is too small that's can't get more profit for us, because 14-25 year old is younger we should to open more customer arrange Why do you think the range is too small. It is good to have a range and target on this group of people with special designed products. If you think there is a need to increase the age range, can you suggest how? There is a need to understand the potential of the products for different age group such as baby food or food for elder adults. Also, is there facility for the study and invention.
2	Problems identified without valid reason	Market analysis, you can give more examples on that. Revise on reason 3, you can argue that the target customer is different in two companies. You can try to write in this way: Since the target group is different, there is a need to understand the needs of the target group before any decision is made.
1	Irrelevant problems identified (subject to students writing work review)	'The snack more delicious and more attractive,' how can you make the snack more disciuos and more attractive?

Appendix G Coding examples for students' argumentative writing

Mark	Student writing work	Students' Example
5	All reasons are valid complete with detailed explanation	I've read the information of your company, but I'm not sure whether it's good or not to invest your company. There are something more that I would like to know before I make my decision. First, I'm not sure whether Mesa Foods is prior to other similar companies. Although it has strong growth potential and its sales keeps increasing, maybe the total selling amount is way behind others. In that case, I would rather invest other companies which already have steady and better sales. Second, I would like to know your selling strategy for the future. Since we are going to expand your business into a big region. I want to know how you will manage your company and your strategy to sell products in a big area. You need to convince that Mesa Foods will also success in a big region. Last, I am curious about the sales volume of your products except AA Salsa. I would like to know how much profits are made by those products. If other products didn't sell well, then I hope you can make some changes to improve. Overall, I hope all the products can reach the basic profits. I'm afraid that the current information which you provided is not clear enough please add those information written above and send it to me. We will contract you as soon as the decision was made.
4	Most reasons are valid complete with detailed explanation	About the investing, I know it has the more advantage. But I think it has some question to discuss before making decision. I think buying Mesa Foods is a good idea, and we need more information. Although they enjoyed a 20percent increase in last year and its best-selling product, it maybe be hot at that time. We need to know the company's 經營方針(translation: marketing strategies). For example, like as marketing 分析(translation: analysis) and costume and so on. Our company make target customer 14-25 year olds. The age group like to try new thing, but it doesn't mean that they love snack food. Although AA salsa is best-selling product, we can't understand the age group is more than other age groups. If Omni wants to buy it, it maybe try to do the other flavours. Mesa enjoyed a 20% small region of the country, but Omni wants to focus on the country. The two company set the different direction. It's difficult to do. Maybe Omni need to do survey throughout the country. Finally, to buy the company, we need to know more information.
3	At least 1 reason valid complete with detailed	I just read about the passage from strategic planning department on buying mesa foods and marketing their popular snacks. I know you feel strongly about this investment, but you really understand this company's anything? I don't think so because

	explanation	<p>you have to more analysed the various factors. First, mesa enjoyed a 20 percent increase in profits last year, we don't know how many percent increase in profits by other products. We need to be clear about this part. Secondly, AA Salsa, has had increased sales over each of the past three years. We don't know how many percent in three years has had increased sales. And three years means constantly years? We should figure out this part. Thirdly, since Omni Inc. is interested in reaching 14-25 year olds, the age group that consumes the most snacks food. By comparing these similar product from the same business. This age group really like snacks produce by this company? Therefore, we don't know more problem with this company. So I disagreement to the investment this company. If you insist on invest, you have to consider more problem again.</p>
2	At least 1 reason without detailed explanation	<p>I just read about the passage from strategic planning department on buying mesa foods and marketing their popular snacks. I know you feel strongly about this investment, but I will not agree the plan. First, mesa Foods has been succeeded in a small region, but it would not success in a big region. We need do the marketing test for mesa Foods's product in the big region. Secondly, our target customer is 14-25 years old, and we can't sure 14-25 years old would like AA salsa that yours popular products. Thirdly, mesa foods only showed the profit of AA salsa, but the other products might has low profit. If we buy the mesa foods, it is big risk in this case. Therefore, by taking all these into consideration, it is wise to reconsider the investment.</p>
1	All reasons are invalid	<p>Hi, my name is XX. I have finished to reading the e-mail about the memorandum issued by the strategic department at Omni Inc. But I have three reasons about why I will disagreement to the investment. We don't have correct information to know how many all the profits to shave this 20%. One-fifth of the profits from AA salsa. But have not clearly explain of other four-fifths profits. If the AA salsa will not popular. That the company will reduce 20% benefit. I think the company should have other popular products. And the consume group are not only between 14-25 year olds. 14-25 years olds group are not the main customer who have ability to buy. They may need to buy other peoples' help. I think that consumers shouldn't only 14-25 young people. On the basis of this reason, I think we should not buy at this time. We buy it when the timing is good.</p>

Appendix H Transcription Conventions

A full discussion of CA transcription notation is available in Atkinson and Heritage (1984). Punctuation marks are used to capture characteristics of speech delivery, but not to mark grammatical units.

[indicates the point of overlap onset
]	indicates the point of overlap termination
=	a) turn continues below, at the next identical symbol b) if inserted at the end of one speaker's turn and at the beginning of the next speaker's adjacent turn, it indicates that there is no gap at all between the two turns
(3.2)	an interval between utterances (3 seconds and 2 tenths in this case)
(.)	a very short untimed pause
e:r	the::: indicates lengthening of the preceding sound
-	a single dash indicates an abrupt cut-off
?	rising intonation, not necessarily a question
!	an animated or emphatic tone
,	a comma indicates low-rising intonation, suggesting continuation
.	a full stop (period) indicates falling (final) intonation
CAPITALS	especially loud sounds relative to surrounding talk
° °	utterances between degree signs are noticeably quieter than surrounding talk
↑ ↓	indicate marked shifts into higher or lower pitch in the utterance following the arrow
> <	indicate that the talk they surround is produced more quickly than

	neighbouring talk
.hh	speaker in-breath
hh	speaker out-breath
HA HA heh heh	laughter transcribed as it sounds
→	arrows in the left margin pick out features of especial interest

Additional symbols

Dui <i>yes</i>	English translation is italicised, and follow Chinese pinyin in next line.
((action))	indicate an action or non-verbal expression made by the speaker
S1	student abbreviation

BIBLIOGRAPHY

- Ajzen, I., 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), pp. 179–211.
- Al-Hazmi, S. H. & Scholfield, P., 2007. Enforced revision with checklist and peer feedback in EFL writing: the example of Saudi university students. *Scientific Journal of King Faisal University (Humanities and Management Sciences)*, 8(2), pp. 237-267.
- Allwright, D., 1991. *Focus on the Language Classroom, An Introduction to Classroom Research for Language Teachers*. Cambridge: Cambridge University Press.
- Amores, M., 1997. A new perspective on peer-editing. *Foreign Language Annals* , 30, pp. 513–23.
- The Association for Educational Communications and Technology, 2001. *Research Handbook*. [Online] Available at:
http://learnngen.org/~aust/EdTecheBooks/AECT_HANDBOOK96/41/41-01.html
[Accessed 25 08 2013].
- Atkinson, J. & Heritage, J., 1984. *Structures of social action: studies in conversation analysis*. Cambridge: Cambridge University Press.
- Barbalet, J., 2009. Pragmatism and Symbolic Interactionism. In: *The New Blackwell Companion to Social Theory*. Blackwell Publishing Ltd., pp. 199-217.
- Barnes, D., 2008. *Exploring Talk in Schools*. [Online] Available at:
http://www.corwin.com/upm-data/23512_01_Mercer_Ch_01.pdf [Accessed 13 08 2013].
- Berg, B. L., 2007. *Qualitative Research Methods for the Social Sciences*. 6 ed. San Francisco: Pearson.
- Berg, I., Admiraal, W. & Pilot, A., 2006. Designing student peer assessment in higher education: analysis of written and oral peer feedback. *Teaching in Higher Education*, 11(2), pp.135-147.
- Bernstein, B., 1996. *Pedagogy, Symbolic Control and Identity*. London: Taylor & Francis.
- Black, P. & Wiliam, D., 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), pp. 7-74.
- Blisshen, E., 1969. *The School That I'd Like*. Harmondsworth: Penguin.
- Bruffee, K. A., 1984. Collaborative learning and the 'conversation of mankind'. *College English*, 46(7), pp. 635-652.
- Bryman, A., 2007. Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*, 1(1), pp. 8-22.

- Bryman, A., 2008. *Social Research Methods*. New York: Oxford University Press.
- Carr, L. T., 1994. The strengths and weaknesses of quantitative and qualitative research: what method for nursing?. *Journal of Advanced Nursing*, 20(4), pp. 716-721.
- Chou, M., 1998. How peer negotiations shape revision. In: J. Katchen & Y. Leung, eds. *The Proceedings of the Seventh International Symposium on English Teaching*. Taipei: Crane Publishing Co., pp. 349-359.
- Connor, U. & Asenavage, K., 1994. Peer response groups in ESL writing classes: how much impact on revision?. *Journal of Second Language*, 3, pp. 257-276.
- Cotteral, S. & Cohen, R., 2003. Scaffolding for second language writers: producing an academic essay. *ELT Journal*, 57(2), pp. 158-166.
- Deakin, R., McCombs B., Haddon A., Broadfoot P. & Tew, M., 2007. The ecology of learning: factors contributing to learner-centered classroom cultures. *Research Papers in Education*, 22, pp. 267-307.
- Dobao, A. F., 2012. Collaborative dialogue in learner–learner and learner–native speaker interaction. *Applied Linguistics*, 33(3), pp. 229-256.
- Dunn, K. E. & W. Mulvenon, S., 2009. *A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative*, s.l.: Practical Assessment, Research & Evaluation.
- Falchikov, N., 2001. *Learning Together: Peer Tutoring in Higher Education*. London: Routledge.
- Falchikov, N., 2004. *Improving Assessment Through Student Involvement: Practical Solutions for Aiding Learning*. London: Routledge.
- Falchikov, N. & Goldfinch, J., 2000. Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, pp. 287-322 .
- Field, A. P., 2000. *Discovering Statistics Using SPSS for Windows: Advanced Techniques for Beginners*. London: Sage Publications.
- Fontana, D. & Fernandes, M., 1994. Improvements in mathematics performance as a consequence of. *British Journal of Educational Psychology*, 64 , pp. 407-417.
- Gielen, S. Peeters, E., Dochy, F., Onghena & P. Struyven, K., 2010. Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20, pp. 304-315.
- Greene, J. C., Caracelli , V. J. & Graham , W. F., 1989. Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), pp. 255-274.
- Guerrero, M. C. M. de. & Villamil, O.S., 1994.. Social-cognitive dimensions of interaction in L2 peer revisions. *Modern Language Journal*, 78(4), pp. 484-496.

- Guerrero, M. C. M. de. & Villamil, O. S., 2000. Activating the ZPD: mutual scaffolding in L2 peer revision. *The Modern Language Journal*, 84(i), pp. 51-68.
- Hedgcock, J. & Lefkowitz, N., 1992. Collaborative oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing*, 1(3), pp. 255-276.
- Hirvela, A., 1999. Collaborative writing instruction and communities of readers and writers. *TESOL Journal*, 8(2), pp. 7-12.
- Howe, K. R., 1988. Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational researcher*, 17, pp. 10-16.
- Hu, G., 2005. Using peer review with Chinese ESL student writers. *Language Teaching Research*, 9(3), pp. 321-342.
- Hu, G. & Lam, S., 2010. Issues of cultural appropriateness and pedagogical efficacy: exploring peer review in a second language writing class. *Instr Sci*, 38, pp. 371-394.
- Hughes, G. B., 2009. Formative assessment practices that maximize learning for students at risk. In: *Handbook of Formative Assessment*. Hoboken: Routledge, pp. 212-232.
- Jacobs, G. M., Curtis, A., Braine, G., & Huang, S-Y., 1998. Feedback on student writing: taking the middle path.. *Journal of Second Language Writing*., 7, pp. 307-317.
- Johnson, R. B. & Onwuegbuzie, A. J., 2004. Mixed method and mixed model research. In: B. Johnson & L. Christensen, eds. *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Boston: MA: Allyn and Bacon, pp. 408-431.
- Kamimura, T., 2006. Effects of peer feedback on EFL student writers at different levels of english proficiency: a Japanese context. *TESL CANADA Journal*, pp. 12-39.
- Kane, J. S. & Lawler, E. E. I., 1978. Methods of peer assessment. *Psychological Bulletin*, 85(3), pp. 555-586.
- Krathwohl, D. R., 1993. *Methods of Educational and Social Science Research: An Integrated Approach*. White Plains, NY: Longman.
- Kubany, A. J., 1957. Use of sociometric peer nominations in medical education research. *Journal of Applied Psychology*, 41(6), pp. 389-394.
- Kwok, L., 2008. Students' perceptions of peer evaluation and teachers' role in seminar discussions. *Electronic Journal of Foreign Language Teaching*, 5(1), pp. 84-97.
- Leki, I., 1990a. Potential problems with peer responding in ESL writing classes. *CATESOL Journal*, 3, pp. 5-17.
- Leki, I., 1990b. Coaching from the margins: issues in written response.. In: B. Kroll, ed. *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press, pp. 57-68.

- Leki, I., 1991. The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals*, 24, pp. 203–218.
- Leung, C. & Mohan, B., 2004. Teacher formative assessment and talk in classroom contexts: assessment as discourse and assessment of discourse. *Language Testing*, 21, pp. 335-359.
- Lin, S. S. J., Liu, E. Z. F. & Yuan, S. M., 2002. Student attitudes toward networked peer assessment: case studies of undergraduate students and senior high school students. *International Journal of Instructional Media*, 29, pp. 241–256.
- Liu, J. , 1997. A comparative study of ESL students' pre-/post-conceptualisations of peer review in L2 composition, In: Hyland, K. & Hyland, F., 2006. Feedback on second language students' writing. *Language Teaching*, 39(2), pp. 83-101.
- Liu, J. & Hansen, J. G., 2002. *Peer Response in Second Language Writing Classrooms*. The University of Michigan Press.
- Liu, J. & Sadler, R. W., 2003. The effect and affect of peer review in electronic. *Journal of English for Academic Purposes*, 2, pp. 193-227.
- Lockhart, C. & Ng, P., 1993. How useful is peer response?. *CATESOL Journal*, 5(1), pp. 5-17.
- Loddington, S., 2008. Peer assessment of group work: a review of the literature. [Online] Available at: webpaproject.lboro.ac.uk/files/WebPA_Literature%20review%20.pdf [Accessed 13 08 2012].
- Long, M. H., 1996. The role of the linguistic environment in second language acquisition. In: W. C. Ritchie & T. K. Bhatia, eds. *Handbook of Research on Language Acquisition: Second Language Acquisition. Vol. 2.* New York: Academic Press, pp. 413–468.
- Long, M. H. & Porter, P. A., 1985. Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly*, 19(2), pp. 207-228.
- Lumley, T., Brown, A., Hinkel, E., Lumley, T. & Brown, A., 2005. Research methods in language testing. In: E. Hinkel, (ed). *Handbook of Research in Second Language Teaching and Learning*. Mahwah, NJ: Lawrence Erlbaum, p. 833—855.
- Mackey, D., 2006. *Send Me A Message*. 1st ed. New York: McGraw-Hill.
- Mangelsdorf, K. & Schlumberger, A., 1992. ESL student response stances in a peer-review task. *Journal of Second Language Writing*, 1(3), pp. 235-247.
- Martinez, J. G. R. & Martinez, N. C., 1992. Re-examining repeated testing and teacher effects in a remedial mathematics course. *British Journal of Educational Psychology*, 62, pp. 356-363.

- McDowell, L., 1995. The impact of innovative assessment on student learning. *Innovation in Education and Training International*, 32, pp. 302–313.
- McGroarty, M. E. & Zhu, W., 1997. Triangulation in classroom research: a study of peer revision. *Language Learning*, 47(1), pp. 1–43.
- Mendonça, C. O. & Johnson, K. E., 1994. Peer review negotiations: revision activities in ESL writing instruction. *TESOL Quarterly*, 28, pp. 745–769.
- Mercer, N. & Hodgkinson, S. 2008. *Exploring Talk in School: Inspired by the Work of Douglas Barnes*. London: Sage
- Min, H. T., 2003. Why peer comments fail?. *English Teaching and Learning*, 27(3), pp. 85–103.
- Min, H. T., 2005. Training students to become successful peer reviewers. *System*, 33(2), pp. 293–308.
- Min, H. T., 2006. The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15, pp. 118–141.
- Min, H. T., 2008. Reviewer stances and writer perceptions in EFL peer review training. *English for Specific Purposes*, 27, pp. 285–305.
- Morais, A. M., 2002. Basil Bernstein at the micro level of the classroom. *British Journal of Sociology of Education*, 23(4), pp. 559–569.
- Nagel, T., 1986. *The View from Nowhere*. New York: Oxford University Press.
- Nelson, G. L. & Carson, J. G., 1998. ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing*, 7(2), pp. 113–131.
- Nelson, G. & Murphy, J., 1992. An L2 writing group: task and social dimensions. *Journal of Second Language Writing*, 1, pp. 171–93.
- Nelson, G., & Murphy, J., 1993. Peer response groups: do L2 writers use peer comments in revising their drafts?. *TESOL Quarterly*, 27, pp. 135–142.
- Nicol, D. & Macfarlane, D., 2006. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), pp. 199–218.
- Paulus, T., 1999. The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8, pp. 265–289.
- Pryor, J. & Crossouard, B., 2008. A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), pp. 1–20.
- Pryor, J. & Crossouard, B., 2010. Challenging formative assessment: disciplinary spaces and identities. *Assessment & Evaluation in Higher Education*, 35(3), pp. 265–276.

- Race, P., 2001. A briefing on self, peer and group assessment. *LTSN Generic*, 9.
- Rollinson, P., 2005. Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), pp. 23-30.
- Sadler, D., 1989. Formative assessment and the design of instructional systems. *Instructional Science*, 18, pp. 145-165.
- Scheraga, M., 1999. *Practical English Writing Skills: A Handbook with Practice*. 2nd ed. Lincolnwood, Illinois: Contemporary Books.
- Seedhouse, P., 1995. *L2 Classroom Transcripts: Data in Search of a Methodology?*. [Online]
Available at: http://www.zait.uni-bremen.de/wwwgast/tesl_ej/ej04/a1.html
[Accessed 25 08 2013].
- Sengupta, S., 1998. Peer evaluation: "I am not the teacher". *ELT Journal*, 52 (1), pp. 19-28.
- Singleton, R. A. & Straits, B. C., 2005. *Approaches to Social Research*. 4 ed. New York: Oxford University Press.
- Sluijsmans, D. M. A., Brand-Gruwel, S. & Merriënboer, J. G. v., 2002. Peer assessment training in teacher education: effects on performance and perceptions.. *Assessment & Evaluation in Higher Education*, 27, pp. 443–454.
- Stanley, J., 1992. Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing*, 1, pp. 217-233.
- Swann, J., 1994. Researching language and literacy in social context. In: J. Maybin & B. Stierer, eds. *Observing and Recording Talk in Educational Settings*. Clevedon: Multilingual Matters, pp. 26-48.
- Tashakkori, A. & Teddlie, C., 2003. *Handbook of Mixed Methods in Social and Behavioral Research*.. Thousand Oaks, CA: Sage Publications.
- Topping, K., 1998. Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, pp. 249–276.
- Topping, K., 2003. Self and peer assessment in school and university: reliability, validity and utility. *Innovation and Change in Professional Education*, 1, pp. 55-87.
- Topping, K., 2009. Peer assessment. *Theory into Practice*, 48(1), pp. 20-27.
- Topping, K., Smith, E.F. , Swanson, I. & Elliot, A., 2000. Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, 25(2), pp. 149-168.
- Torrance, H. & Pryor, J., 1998. *Investigating Formative Assessment*. 1st ed. Philadelphia: Open University Press.

- Tsui, A. & Ng, M., 2000. Do secondary L2 writers benefit from peer comments?. *Journal of Second Language Acquisition*, 9(2), pp. 147-170.
- Varonis, E. & Gass, S., 1985. Non-native conversations: A model for NfM. *Applied Linguistics*, 6, pp. 71-90.
- Villamil, O. & Guerrero, M.C.M. de, 1996. Peer revision in the L2 classroom: social-cognitive activities, mediating strategies, and aspects of social behavior. *Journal of Second Language Writing*, 5, pp. 51-75.
- Villamil, O. & Guerrero, M. C. M de, 1998. Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, 19, pp. 491-514.
- Walsh, S., 2011. *Exploring Classroom Discourse: Language in Action*. s.l.:Routledge.
- Weaver, R. L. & Cottrell, H. W., 1989. Motivating students: stimulating and sustaining student effort sustaining student effort. *College Student Journal*, 22(1), pp. 22-32.
- Wen, M. L. & Tsai, C.-C., 2006. University students' perceptions of and attitudes toward (Online) peer assessment. *Higher Education*, 51(1), pp. 27-44.
- Whiting, B., Van Burgh, J. W. & Render, G. F., 1995. *Mastery of Learning in the Classroom*. San Francisco, CA.
- Yang, M., Badger, R. & Yu, Z., 2006. A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, Volume 15, pp. 179-200.
- Zhang, S. Q., 1995. Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, 4, p. 209-222.
- Zhang, S. Q., 1999. Thoughts on some recent evidence concerning the affective advantage of peer feedback. *Journal of Second Language Writing*, 8, p. 321-326.
- Zhu, W., 1995. Effects of training for peer response on students' comments and interaction. *Written Communication*, 1(4), pp. 492-528.